DOCUMENT RESUME

ED 373 098 TM 021 993

TITLE	Charting the Course toward Instructionally Sound Assessment. A Report of the Alternative Assessment Pilot Project.
INSTITUTION	California Assessment Collaborative, San Francisco.
SPONS AGENCY	California State Dept. of Education, Sacramento.
PUB DATE	Sep 93
CONTRACT	8363
NOTE	139p.
AVAILABLE FROM	California Assessment Collaborative, 730 Harrison Street, San Francisco, CA 94107.
PUB TYPE	Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS	*Cooperation; *Educational Assessment; Educational Change; Elementary Secondary Education; Information Dissemination; *Pilot Projects; Research Methodology; *School Districts; Special Needs Students; Standardized Tests; *Test Construction; Test Use
IDENTIFIERS	*Alternative Assessment; *California; Collaborative Research; Performance Based Evaluation

ABSTRACT

The California Assessment Collaborative (CAC) was designed as a three-year effort to identify, validate, and disseminate alternatives to standardized testing throughout the state. This report highlights the results from the first year and comments on the work of 22 diverse pilot projects that serve as examples of site-based development of performance assessments. The project must address all subject and content areas in at least two grade levels and pay particular attention to issues of assessment for students with special needs. They must pay close attention to issues and practices of assessment development at the state and national levels and must demonstrate a potential for collaboration with researchers, publishers, and other partners. Although, at the end of this first year, none of the project models is ready for dissemination, each has already contributed to the understanding of systemic change at school and district levels. Costs and the impacts of these projects are also examined. Three appendixes contain a discussion of methodology and lists of contact persons and credits. Study findings are presented in 26 figures. (Contains 31 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TM

ED 373 098

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

"PERMISSION TO REPRODUCE THIS MATERIAL IN **OTHER THAN PAPER COPY** HAS BEEN GRANTED BY

T. SMITH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CAC
CALIFORNIA ASSESSMENT
COLLABORATIVE

Charting the Course

Toward Instructionally Sound Assessment

TM021993

BEST COPY AVAILABLE

CHARTING THE COURSE TOWARD INSTRUCTIONALLY SOUND ASSESSMENT

A Report of the Alternative Assessment Pilot Project

Pursuant to Chapter 5
(Section 60730 of the California Education Code)

California Assessment Collaborative
Catherine Jamentz, Director

September 1993

A collaborative project among the Greater ACE Consortium, the Greater San Diego Area Consortium, and Far West Laboratory for Educational Research and Development

ACE Consortium

SDCS
SAN DIEGO CITY SCHOOLS

SAN DIEGO COUNTY OFFICE OF EDUCATION

FarWest LABORATORY

PRINTED WITH SOY INK

This report is available through the California Assessment Collaborative,
730 Harrison Street, San Francisco, CA 94107. Phone (415)241-2704.

This document is supported by state funds made available by California Department of Education contract number 8363. The Alternative Assessment Pilot Project was mandated by California Assembly Bill 40, 1990, Quackenbush (Education Code Section 60730). The contents of this report do not necessarily reflect the views or policies of the California Department of Education.

ACKNOWLEDGMENTS

We wish to express our appreciation to the hundreds of teacher participants in the California Assessment Collaborative's pilot projects. The risks they have taken, their struggles and successes, have charted our course.

We also wish to thank Dale Carlson and Sue Bennett of the California Learning Assessment System office in the California Department of Education for their support of this project and their leadership in assessment reform.

About the Partners in the California Assessment Collaborative

ACE Consortium — The Assessment Curriculum and Evaluation (ACE) Consortium is housed at the Santa Clara County Office of Education. This consortium of approximately 50 school districts and county offices was founded in 1978 for the purpose of working together on projects of mutual interest.

Greater San Diego Area Consortium — The consortium is made up of the San Diego City Schools and districts throughout San Diego County. San Diego City and County Schools have initiated and supported many alternative assessment efforts.

Far West Laboratory (FWL) — Far West Laboratory for Educational Research and Development is located in San Francisco. It conducts research, provides technical assistance and training and offers evaluation support to schools and education agencies.

Dear Reader,

"Charting the Course Toward Instructionally Sound Assessment" is the report of the California Assessment Collaborative (CAC), the California Department of Education's Alternative Assessment Pilot Project. Conceived by Assembly Bill 40 (Quackenbush), the CAC supports and studies the development of performance assessment strategies in schools and districts throughout the state. The project seeks to amplify the voice of teachers in the development of assessments at the local, state and national level and to work with teachers to strengthen their assessment skills.

The CAC is a collaboration between two consortia of school districts, the Greater Assessment Curriculum Evaluation (ACE) Consortium in the San Francisco Bay Area and the Greater San Diego Area Consortium, which includes San Diego City and County Schools. Far West Laboratory for Educational Research and Development provides technical assistance and conducts research on the CAC's pilot projects.

The CAC was designed as a three-year effort to systematically identify, validate, and disseminate alternatives to standardized testing throughout the state. This report represents findings from our first year of work. In it we have highlighted the work of 22 diverse pilot projects which serve as examples of site-based performance assessment development efforts and which span the requirements of AB40 for projects which

- address all subject areas/content areas in at least two grade levels,
- pay particular attention to issues of assessment for students with special needs,
- pay close attention to issues and practices of assessment development at the state and national level, and
- demonstrate the potential of collaboration with researchers, publishers and other collaborative partners.

Collectively, these projects have taught us a great deal. Although, as of the end of the first year, none provide assessment models ready for dissemination, each has revealed some promising practices and pitfalls in the site-level development of assessment alternatives and has contributed to our understanding of systemic change efforts at school and district levels. The projects represent the complex planning, resource allocation, professional development, and implementation strategies which are an essential part of assessment reform.

This report will also examine the costs and impact of these performance assessment efforts and build on all these findings to make recommendations for the direction of future work.

This report is written to several audiences:

- The California Department of Education and its collaborative partners, who have set out to develop the new California Learning Assessment System
- State legislators and other policy makers who must consider the costs and benefits of continued performance assessment development
- School practitioners who are interested in or are considering their own performance assessment development efforts

We welcome other readers from the field of assessment and the academic and research communities.

Our work is by no means complete. Our future work will likely focus on continued development and validation of the most promising projects, continued study of issues of implementation and dissemination, and further dissemination efforts.

Writing Committee
California Assessment Collaborative

TABLE OF FIGURES

Figure 1 General Learner Outcomes
Berryessa Union School District 14

Figure 2 Student Outcomes
Palo Alto Unified School District 15

Figure 3 Essential Questions
ACE K-2 Alternative Assessment Project 17

Figure 4 Grade 8 Humanities Course Outline
Muirlands Middle School
San Diego City Schools 19

Figure 5 Letter to Parents
Stevens Creek School
Cupertino Union School District 20

Figure 6 Characteristics of Authentic Assessments 29

Figure 7 Portfolio Contents
O'Farrell Community School
San Diego City Schools 30

Figure 8 Progressive Revisions to Open-Ended Science Questions
Chula Vista Elementary School District 31

Figure 9 Interpretation Guide
ACE K-2 Alternative Assessment Project 33

Figure 10 Progressive Revisions to a Math Scoring Rubric
Mt. Diablo Unified School District 34

Figure 11 Initial Draft of Rubric for Narrative Writing
Stevens Creek School
Cupertino Union School District 35

Figure 12 Reading Scale from the California Learning Record 36

Figure 13 Prototype Scoring Guide
Redwood City School District 37

Figure 14 Grade 2 Student Progress Report
La Mesa — Spring Valley School District 41

Figure 15 GATE Nomination Form
Oceanside Unified School District 45

Figure 16 Essential Questions and Indicators
ACE K-2 Alternative Assessment Project 47

Figure 17 Scoring Guide
Chula Vista Elementary School District 48

Figure 18 Student Portfolio
Newcomer High School
San Francisco Unified School District 50

Figure 19 Performance Prompt — Growing Up
ACE K-2 Alternative Assessment Project 55

Figure 20 Project Criteria — Presenting Myself
O'Farrell Community School
San Diego City Schools 56

Figure 21 Writing Folder Guidelines
La Mesa — Spring Valley School District 57

Figure 22 Questions Compiled from Student Reflection Forms 59

Figure 23 Final Portfolio Check-Off Sheet
Coordinated Science
Monte Vista High School
Grossmont Union High School District 60

Figure 24 Teacher, Parent/Guardian, Peer
Portfolio Evaluation Form
O'Farrell Community School
San Diego City Schools 61

Figure 25 Assessment Vision and Policy
San Diego City Schools 64

Figure 26 Scaffolded Performance Prompt
ACE K-2 Alternative Assessment Project 66

CONTENTS

Chapter 1 — Introduction 1

Chapter 2 — A Conceptual Map 11

- Articulating Content Standards 13
- Developing Meaningful and Fair Assessments 22
- Developing Teacher Capacity to Use Assessment to Improve Instruction 44
- Developing Student Capacity to Use Assessment to Improve Learning 54
- Determining and Monitoring Consequences 63

Chapter 3 — First Person Narratives 69

- Discovering Teacher and Student Capacity: The Story of the ACE K-2 Project 73
- Creating Opportunities for Students to Learn and Practice Self-Assessment: The Monte Vista High School Science Portfolio Assessment Project 76
- A Place Where Dreams Happen: O'Farrell Community School 78
- Developing and Using Meaningful Tasks to Improve Learning: Chula Vista Elementary School District 81
- Building a District Culture of Collegial Inquiry: Hillsborough School District 83
- All Students Will Succeed: The Story of Redwood City School District 85
- Overview of Issues: A Superintendent's Perspective on Assessment Development 88

Chapter 4 — Analyzing Benefits and Costs 91

- The Benefits and Costs of Traditional Assessments 95
- A Framework for Analysis of the Benefits of Performance Assessments 98
- A Framework for Analysis of the Costs of Performance Assessments 105
- Putting the Cost/Benefit Frameworks to Use 112

Chapter 5 — Conclusions and Implications 113

References 121

Appendices

- A: Research Methodology 127
- B: CAC Pilot Project Contacts 133
- C: Credits 135

INTRODUCTION

California Defines its Vision

Over the course of the last ten years, the California Department of Education has detailed a comprehensive vision for school reform. Initiated by the publication of a series of innovative curriculum frameworks and further brought into focus by the reports of the Child Development, Elementary Grades, Middle and High School Task Forces, this vision challenges educators, policy makers, and community leaders to rethink traditional goals, structures, and instructional strategies in our schools.

> *Our schools were originally designed to meet the needs of a rural, agrarian society. Gradually, they adapted to the needs of an urban, industrial society. Now, high schools must enter the next stage of improvement activities and ... meet the challenges of an information- and knowledge-based, global society.*
>
> *As we approach the twenty-first century, more students than ever before need to be educated to higher levels. ... Schools must prepare students who know how to think and to learn; who know how to gather, organize and analyze information and then apply it to solving a problem; and who can work collaboratively with others.*
> (California High School Task Force, 1992)

> *The service and technology-driven global economy of the twenty-first century will demand a much higher level of intellectual attainment for a far greater percentage of workers than did the smokestack and agriculture-dominated one of the past. American business leaders have been telling educators that they seek employees who not only can read and compute but also can think creatively and critically and adapt to change.*
> (Elementary Grades Task Force Report, 1992)

As the demands of society have changed, so too has our understanding of the learning process. In "Toward The Thinking Curriculum" Resnick and Klopfer (1989) summarize findings from recent advances in cognitive psychology: "To know something is not just to have received information, but also to have interpreted it and related it to other knowledge." That is, the learner constructs meaning by linking new information to prior knowledge and experience.

A view of students as active participants in the learning process has major implications for curriculum design and instructional practice:

> *The traditional grade school curriculum consisted of a hierarchical sequence of basic academic skills that students were expected to acquire, often through dogged practice. This approach reflected a view of children as empty vessels that the teacher filled up with knowledge by pouring in an agreed-on inventory of skills in an agreed-on order (from simple to complex, with the most complex "skills" of thinking and problem solving reserved for the later years of education). But modern cognitive research has found that ... all children are quite capable of sophisticated thought processes from the beginning of their formal education. Indeed, they truly learn only when they are afforded the opportunity to actively incorporate what they are studying into their own experiences, concepts and understandings of how the world works.*
> (Elementary Grades Task Force Report, 1992)

The California curriculum frameworks and school level task force reports call for instructional practices which engage all students in the same high level "thinking curriculum" — providing opportunities that cause students to formulate and solve problems, criticize their own work, work in teams, and communicate about what they are learning. Teachers are envisioned as coaches:

> *Coaching means that every student learns in an active, participatory context rather than a passive one; contributes; is encouraged to succeed; and can excel if expectations are kept high. A teacher as coach conceptualizes the job as making sure that all students learn the curriculum, instead of thinking of the job as presenting the material and leaving it up to the students to understand it.*
> (California High School Task Force, 1992)

The Pivotal Role of Assessment

The keepers of the vision for California's schools recognize the pivotal role of assessment in bringing about necessary changes in curriculum and instructional practice at all levels of schooling.

> *A basic rule of education is that 'What you test is what you get.' . . . The problem with many assessment instruments used at the elementary school level is that they still reinforce the old skills-based curriculum. Multiple-choice tests that require students to fill in bubbles on machine-graded answer sheets have tended to put a premium on rudimentary academic skills. ... At the same time teachers have not been spurred by such tests to foster their students' capacities to perform the generative acts of analysis, organization, communication, creativity and other higher order thinking processes that are the hallmark of a rich, meaning-centered curriculum.*
> (Elementary Grades Task Force Report, 1992)

> *Current assessment programs are inadequate. New measurement practices must be developed which encompass an expanded range of educational outcomes.*
> (Middle Grade Task Force, 1987)

> *Assessment, at all levels, is seen as the key strategy in bringing about significant educational reform. Moving beyond the standardized, multiple choice test as the primary accountability tool, [performance-based assessments] help educators measure what matters — including a student's ability to analyze, organize, interpret, explain, synthesize, evaluate, and communicate important experiences.*
> (California High School Task Force, 1992)

In 1989 the California Assessment Program (CAP), recently redesignated the California Learning Assessment System (CLAS), introduced a vision of assessment designed to support comprehensive school reform. Long a leader in innovative strategies for assessing student writing, CAP sponsored two "Beyond the Bubble" conferences statewide and announced plans to pilot performance assessments.

Performance assessments require students to tackle complex tasks involving many possible solutions. Innovative assessment strategies engage students in collaborative activities, investigations, or hands-on problem solving activities similar to those they might experience in the world outside school. Students are challenged to keep portfolios of their work over time and to reflect on, and take increased responsibility for, improving their work.

In October 1991, Governor Pete Wilson signed a major piece of school reform legislation establishing plans for a new state student assessment system. The legislation (SB662), sponsored by Senator Gary Hart, provided for the assessment of all students in grades 4, 8, and 10 against common statewide performance standards. The bill called for the development of a system designed to "make assessment an integral part of the instructional process" and "to facilitate the development of each and every pupil to become a self-motivated, competent, life-long learner."

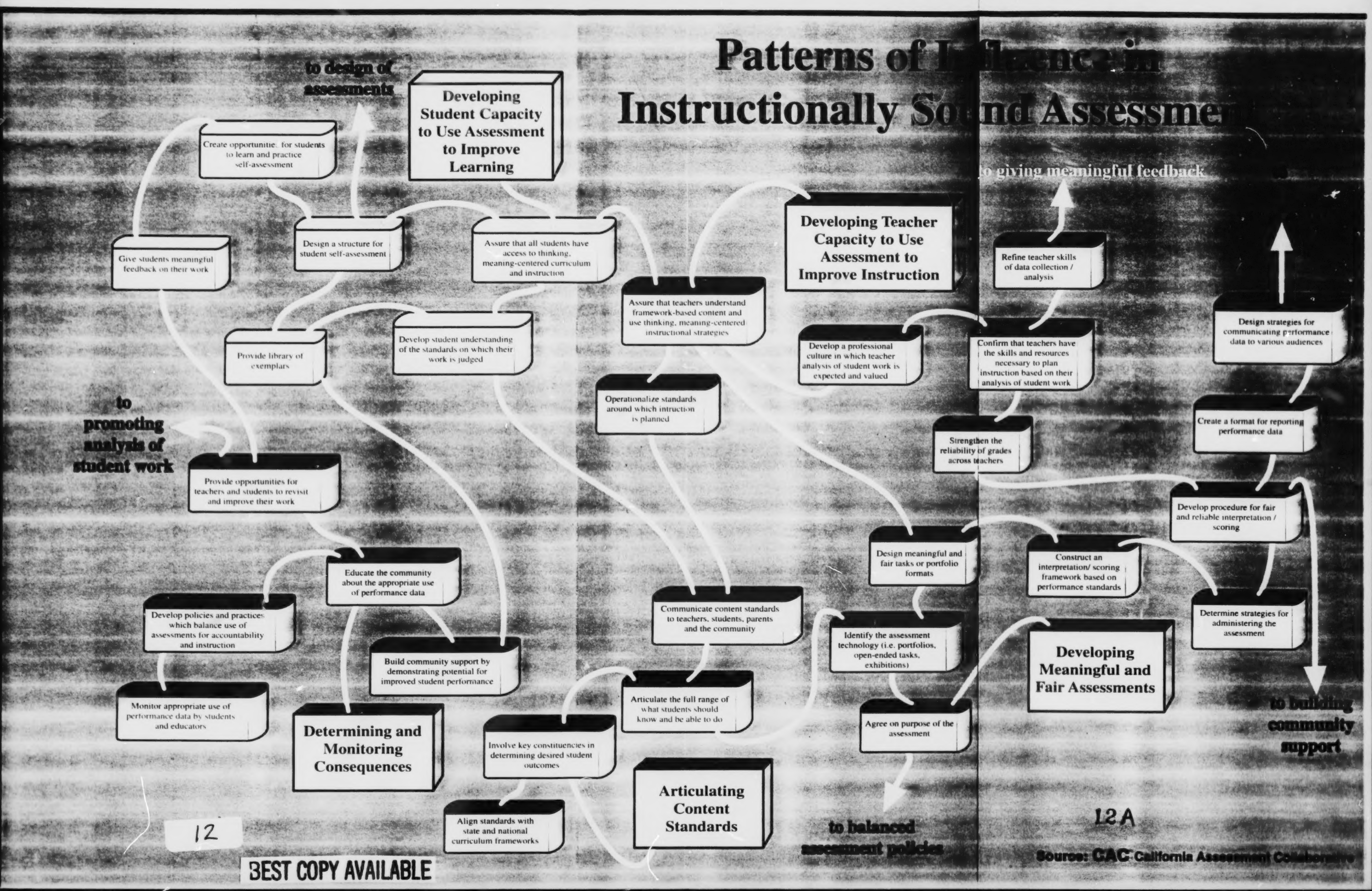
Soon after its passage, CLAS submitted a five year plan for development of the new system which includes performance tasks and portfolios. In January 1993, Governor Wilson requested a $12.5 million augmentation to existing funds for the development of the state assessment system.

Assessment Reform at the Local Level

The need for assessment strategies to support and promote a thinking, meaning-centered curriculum was recognized by many classroom teachers long before the initiative for a statewide system was introduced. Many teachers and other school leaders, often with the support of university researchers, test developers, or professional development experts, began to develop their own assessment tools soon after publication of the first curriculum frameworks. These local efforts sought to generate assessment strategies which would promote a curriculum of richness and depth and provide teachers, students, and parents with data they could use to enhance student performance — assessments which could be called "instructionally sound." In 1990, Assembly Bill AB40 (Quackenbush) provided initial funding for a three-year Alternative Assessment Pilot Project, known as the California Assessment Collaborative (CAC), to support and study these local development efforts.

Between July 1991 and December 1992, the CAC worked with a wide variety of performance assessment development efforts in districts and schools throughout the state and considered strategies for disseminating their work to interested schools. ***Charting the Course Toward Instructionally Sound Assessment*** is the summary of findings from the California Assessment Collaborative's initial work.

Patterns of Influence in Instructionally Sound Assessment



Developing Student Capacity to Use Assessment to Improve Learning

Create opportunities for students to learn and practice self-assessment

Design a structure for student self-assessment

Give students meaningful feedback on their work

Assure that all students have access to thinking, meaning-centered curriculum and instruction

Provide library of exemplars

Develop student understanding of the standards on which their work is judged

Provide opportunities for teachers and students to revisit and improve their work

to design of assessments

to promoting analysis of student work

Developing Teacher Capacity to Use Assessment to Improve Instruction

Refine teacher skills of data collection / analysis

Develop a professional culture in which teacher analysis of student work is expected and valued

Confirm that teachers have the skills and resources necessary to plan instruction based on their analysis of student work

Strengthen the reliability of grades across teachers

to giving meaningful feedback

Articulating Content Standards

Assure that teachers understand framework-based content and use thinking, meaning-centered instructional strategies

Operationalize standards around which intruction is planned

Communicate content standards to teachers, students, parents and the community

Articulate the full range of what students should know and be able to do

Involve key constituencies in determining desired student outcomes

Align standards with state and national curriculum frameworks

Determining and Monitoring Consequences

Educate the community about the appropriate use of performance data

Develop policies and practices which balance use of assessments for accountability and instruction

Build community support by demonstrating potential for improved student performance

Monitor appropriate use of performance data by students and educators

Developing Meaningful and Fair Assessments

Design meaningful and fair tasks or portfolio formats

Identify the assessment technology (i.e. portfolios, open-ended tasks, exhibitions)

Agree on purpose of the assessment

Construct an interpretation/ scoring framework based on performance standards

Develop procedure for fair and reliable interpretation / scoring

Determine strategies for administering the assessment

Create a format for reporting performance data

Design strategies for communicating performance data to various audiences

to balanced assessment policies

to building community support

Source: CAC California Assessment Collaborative

BEST COPY AVAILABLE

The Work of the California Assessment Collaborative in Statewide School Reform

The classroom teacher is central to the vision of assessments which both measure and enhance student performance. Unlike the curriculum-neutral, teacher-proof features of standardized tests, assessments with the potential to improve teaching and learning must be embedded in day-to-day school and classroom practice. The classroom teacher becomes key, not only in devising and administering fair, high quality tasks, but also in interpreting student performance, helping students assess their own progress, and planning instruction based on the results.

The California Assessment Policy Committee for the statewide assessment system, chaired initially by Dr. Thomas W. Payzant, Superintendent of San Diego City Schools, emphasized this centrality in its draft of "Working Principles and Beliefs":

> *One of the most important purposes of the new student assessment system is to nurture local capacity to carry out the more authentic, performance-based assessment at the classroom level — the place where it ultimately matters. In other words, the assessment system and the information it provides must be useful in California's schools and classrooms. Furthermore, if assessment is to realize its full potential to improve instruction, teachers must be central to the process, determining the performance and progress of their students by judging their work. The state must recognize and support the role of teachers as the professional "keepers of the standards."*

Local assessment development and implementation efforts provide a unique opportunity to study assessment reform in the actual context of school life. Through such projects we have opportunities to understand the characteristics of assessments which enhance instructional practice, and the ways teachers, students, and schools use assessment data. Watching and listening to teachers involved in their own assessment development efforts provides clues to the necessary elements and real challenges of an instructionally sound assessment system.

Assuring that assessments enhance and not merely measure student performance means looking beyond the mere design of tasks and scoring processes. Instructionally sound assessments involve not only rich tasks, but the informed participation of teachers and students and the commitment of school leaders and school communities to use assessment data in new ways.

Collectively the CAC pilot projects have illuminated an array of critical and interrelated decision points, design issues, and implementation strategies which have a bearing on the potential of assessment to improve teaching and learning. When considered together, these projects have contributed to the design of a conceptual map defined by five dimensions of practice:

- articulating content standards
- developing meaningful, fair assessments
- building teacher capacity to use assessment to improve instruction
- building student capacity to use assessment to improve learning
- determining and monitoring the consequences of assessment

Each dimension suggests a set of elements or variables which influence the potential of assessment prove "instructionally sound." The preceeding foldout illustrates the conceptual map and the elements associated with each dimension.

The Dimensions of Instructionally Sound Assessment

Determining What Students Should Know and Be Able To Do

In order for assessment to be used to improve teaching and student performance, the standards for what students should know and be able to do must be agreed upon by, and made clear to, both teachers and students. The experience of CAC projects suggests that this goal is more complex than it may at first appear. In practice, the articulation of standards requires decisions about what those standards ought to be, by whom and how they should be determined, and in what ways they will be communicated to others.

Employers in an increasingly technological, information-based society expect workers to have knowledge and skills specific to their field and the ability to apply that knowledge on their own or in groups in situations requiring creative thinking and problem solving. As efforts proceed at the state and national level to determine curriculum standards, local districts and schools consider the alignment of their efforts with this work and seek means for building consensus within the local business, parent, and school community on the full range of knowledge, skills, and capacities they want students to have. The rhetoric of "thinking curriculums," "core curriculums," and "basic skills" is probed as communities struggle to find some common ground on what they really want students to know and be able to do.

In order for standards to contribute to improved teaching and learning, schools must find ways to assure that the standards have meaning to those whose work is founded upon them. Teachers and students work to translate standards into useful tools for planning teaching and learning and for communicating expectations to parents.

Developing Meaningful, Fair Assessments

There are many variables in the design of assessment tools and scoring strategies which influence the degree to which an assessment contributes to improved teaching and learning. Research provides evidence that we get what we assess and not what we don't. Stated differently, assessments communicate what is thought important for teachers to teach and students to learn. Therefore, schools committed to instructionally sound assessment strive to invent assessment tools which require use of the full range of knowledge, skills, and mental habits which are considered important.

School assessment development teams question whether the tasks or strategies they invent are valid measures of the skills and capacities they purport to assess; whether they require students to demonstrate directly their ability to do meaningful work; and whether they are fair and unbiased. At the same time, they struggle with the feasibility of using new assessments in the classroom, or with how their use might require school or curriculum restructuring.

Schools committed to developing instructionally sound assessments seek the contributions of teachers and curriculum specialists who are well grounded in both classroom practice and state-of-the-art knowledge about content and cognition. They design and revise tasks and scoring frameworks after careful study of the student work they generate and carry out rigorous and systematic training of scorers using defined standards of performance.

Building Teacher and Student Capacity To Use Assessment

The availability of meaningful, fair assessment tasks will not assure a positive impact on teaching and learning unless teachers have the capacity to use assessment to improve instruction and students have the capacity to use assessment data to improve their learning. Building capacity means assuring that teachers and students have the motivation, skills, and resources necessary to assess and to use assessment effectively.

As standardized tests have taken on greater importance, teachers have been robbed of, and in some cases, have forfeited, their responsibility for assessment. In such instances, instruction is not based on the assessed needs of students, but on a mandate to cover material or proceed through a scope and sequence of skills. In schools committed to using assessment to improve teaching and learning, teachers take responsibility for determining what students know and what they need to be taught. They have the skills necessary to analyze and judge student work and the time and materials to do it.

In practice, this means that districts and schools build cultures in which the analysis of student work is expected and valued. Teachers have opportunities to explore the essential elements of their discipline and its relation to other disciplines and to learn to observe and analyze student work. Teachers have time to become familiar with standards and exemplars of student performance, and work to increase the reliability of their everyday interpretations and grades. Traditional school structures and schedules which have kept teachers isolated from one another are redesigned to encourage collaboration and consensus-building.

For students to have the capacity to use assessment to improve their learning, they must see an advantage in doing well, have adequate access to high quality curriculum and instruction, and the opportunities and resources necessary to participate in assessment of their own progress.

Schools committed to instructionally sound assessments consider the impact of the consequences of assessment on student motivation to learn. They commit to building school structures which provide all students opportunities to explore essential understandings within and across disciplines and to construct knowledge based on these understandings and personal experience. Students are given opportunities to become familiar with the standards for the scope and quality of their work, have access to a library or catalogue of exemplary performances, are taught the skill of self-assessment, and receive timely and purposeful feedback on their progress.

Determining and Monitoring Consequences

Assessment can enhance or inhibit teaching and learning. It can serve as a roadblock or a compass. The potential of an assessment to enhance teaching and learning is, in part, determined by the consequences of its use. What happens when students and teachers participate in an assessment activity? What are the consequences to those teachers, students, and the school program after assessment results are known?

Instructionally sound assessments are learning experiences. They provide meaningful and timely feedback to teachers and students and are embedded in structures in which teachers and students have opportunities to revisit and improve on their work. They are assessments which encourage, not discourage, teachers and students to learn more.

Schools committed to the use of instructionally sound assessments do not ignore the call for accountability and for assessment data which serve this need. In implementing assessments designed to enhance instruction, schools study the ways in which these data might serve the accountability audience and prioritize the time and resources spent on assessments for each purpose.

In order to assure that assessments enhance teaching and learning, schools take steps to articulate and prioritize the purpose(s) of assessment and determine the needs of its various audience(s). They work to develop accurate and appropriate means of reporting student performance data and coordinate the form and scope of reporting with the needs of each audience. Finally, they constantly monitor both the expected and unintended consequences of assessment for individual students and the school program.

The Contents of this Study

A Conceptual Map

Collectively the CAC's first year pilot projects have illuminated the dimensions and elements of a conceptual map for developing and implementing instructionally sound assessments. Each of the CAC's projects is a work-in-progress. None of them has yet addressed all its elements, and often elements have been addressed with less than satisfactory results. Nonetheless, these projects offer a great deal of information which may be helpful to state and district policy makers responsible for guiding assessment reform and to schools and districts considering their own assessment development work.

Chapter 2 of this report uses the experience of CAC pilot projects to describe strategies and promising practices for addressing the various dimensions of the conceptual map. Examples are provided of the ways in which these projects went about articulating standards, developing tasks, building student and teacher capacity, and determining appropriate consequences. Where possible, excerpts from documents related to the assessment activities are shown, not as exemplars to be copied, but as illustrations of the strategies described. Readers are invited to use these to reflect on, or adapt to, their own work. Each sub-section also includes a set of **"Challenges, Cautions and Caveats,"** warnings from the experience of CAC pilot projects and our research about some of the pitfalls in this work. This section will be most useful to those eager for a point of reference to their own work and willing to adapt the ideas presented here to the unique context of their own district or school. Appendix C of this report gives the names of persons knowledgeable about each project cited should further information be desired.

First Person Narratives

Chapter 3 provides a look at seven individual first person narratives of districts, collaborations among districts, and individual schools involved in assessment work, which are the cultures within which these assessment experiments will root and grow. These deeper looks within a single project help us see, through the eyes of those in the schools, how districts and schools have proceeded with systemic assessment reform efforts. The narratives address questions such as: What factors motivate and sustain assessment reform? What are the impediments? With whom, and how, do teachers and schools leverage support? What can we learn about the pace and timing this reform requires?

The Costs and Benefits of Alternative Assessment

In addition to developing and validating alternative assessments, the California Assessment Collaborative has been charged with analyzing the costs and the benefits of these efforts, in contrast to those of more traditional assessments. **Chapter 4** describes a preliminary examination of these issues.

It is difficult to measure or even accurately estimate the costs of a complex work-in-progress. To do so now would be like estimating the costs of the space program at the point in time when we knew only the cost of building the Wright brothers' plane. Our estimates of benefits at this juncture are similarly limited.

As an initial step toward a more accurate analysis of costs and benefits, the CAC collected data on the uses and costs of traditional assessments, and estimates of the costs and perceived benefits in each of the alternative assessment pilot projects.

It is not necessary to conduct complex studies to reach the conclusion that space travel costs more than traditional forms of locomotion. In this first year of study, the CAC has drawn upon the experience of participants in our pilot projects to determine in what ways the positive impact of alternative assessments might outweigh the costs. We have looked at the effects of this work on curriculum and instruction, on student performance, and on the policies and practices within the school community. Our intent is to use these data to identify specific questions which can be explored in future research.

Preliminary Conclusions and Implications

In **Chapter 5** we share the preliminary conclusions we have drawn from the work of the CAC pilot projects and research. Each conclusion suggests implications for all classroom teachers, the leaders of curriculum, instruction and assessment reform efforts and for district and state-level policy makers. The major conclusions are the following:

1. **The implementation of instructionally sound assessment will require a fundamental redefinition of what it means to teach.**
2. **The development and implementation of instructionally sound assessments must be understood as a complex, systemic process.**
3. **Instructionally sound assessments must root and grow in cultures of collegial inquiry.**
4. **Equity in assessment requires not only attention to the development of fair and unbiased tasks, but to equitable opportunities to learn.**
5. **Assessment development is not just about doing things right, it's about doing the right things.**

Getting to Know the CAC's Phase One Pilot Projects

The CAC's first year pilot projects represent a wide range of assessment development efforts. Assembly Bill 40 requires that projects be identified in at least two grade levels in all major content areas, with special attention to projects which provided opportunities to look at the impact of an assessment on students with special needs.

The projects vary in size and purpose. Some were initiated by individuals or small groups of classroom teachers who set out to invent assessment strategies for the sole purpose of better representing the work their students do. Others seek to create assessment alternatives with the additional purpose of evaluating a program, project, or school. Still others are opportunities to study attempts to influence the development or implementation of pieces of a state or national assessment system headed-up by others.

These projects were selected not as exemplary models or already sophisticated assessment systems, but as examples of local development efforts which had been in place for some length of time. Each is embedded in a district or school which has made a commitment to curriculum reform and improved student performance. For some that commitment has led to significant efforts to redesign school structures and practice. Collectively these projects have taught us a great deal about the issues involved in assessment reform and have helped map our plans for continued development work. The following chart provides a brief description of each pilot project.

	ACE Consortium K-2 Project Berryessa,Burlingame, Evergreen, Hillsborough,Millbrae, Milpitas, Morgan Hill, Oak Grove,Palo Alto, Redwood City, Soquel, and Vallejo School Districts	Teams of teachers working in schools in twelve of the districts in the ACE Consortium have developed a portfolio of interdisciplinary performance tasks. The assessment is used to assist the teacher in instructional planning and to communicate with parents and students about the child's progress.
Greater ACE Consortium Projects	**All Students Will Succeed** Redwood City Elementary All schools in district	This project involves the development of a comprehensive system which includes both a performance-based mathematics and language arts test in grades 2, 5, and 7, developed in collaboration with CTB-McGraw, and eventually student portfolios. The project was designed to provide program evaluation information to schools and the community. As part of the project, the district is developing instructional support materials to better prepare student and teachers for performance assessment.
	Technology-Based Writing Portfolios Stevens Creek School, Cupertino Union School District	Teachers are developing portfolios of student writing in an environment of advanced technological support. This project was initiated as an evaluation tool for the Apple Classroom of Tomorrow program, but has become a tool for teachers to assess student learning and to communicate progress to parents.
	Curriculum-Based Assessment in Mathematics Mt. Diablo Unified All Schools in district	The district has developed its own matrix-sample mathematics test for students in grades three and six, using open-ended math questions. This assessment was initiated for the purpose of program evaluation and to prepare students for CAP.
	Curriculum-Based Performance Assessment San Lorenzo Valley Unified Four elementary schools	Kindergarten through second grade teachers are using performance tasks to assess mathematics, reading, and writing skills. This project was developed in an effort to find an alternative to standardized testing at these grades.
	Curriculum-Based Testing in Special Education San Lorenzo Valley Unified Six elementary and high schools	The special education staff has developed tests in reading fluency, comprehension, writing, and mathematics which are tied to the materials used in the regular classroom. The tests were developed for use in assessing individual student progress and evaluating the Special Education Program.
	Foreign Language Proficiency Palo Alto Unified Two high Schools	Teachers are developing an oral-language proficiency assessment to be used along with a test of writing, reading, and listening skills to determine second language competency.
	HERALD Portfolio Project San Francisco Unified One high school	Teachers at Newcomer High School are developing portfolios to assess the progress of English learners. Portfolios are used by teachers and students to reflect on progress over time and assess the amount and quality of writing in an interdisciplinary curriculum.
	Inquiry Into Alternative Assessment Oakland Unified One Elementary School	Teachers have worked with Educational Testing Service (ETS) to develop and field test performance tasks in mathematics and science in grades K-5. The tasks are used by some teachers as classroom assessment activities.

Greater San Diego Area Consortium Projects - San Diego City Schools	**Elementary Science and Language Arts Assessment Development** San Diego City Schools Seven elementary schools	Elementary teachers are involved in developing and field-testing performance tasks and portfolios in language arts and science. This project was developed to improve instruction and help align instruction to the state frameworks. Teachers use assessment tasks to communicate with students and parents. The project has been supported by the Stuart Foundation for two years.
	Hispanic Reading Project San Diego City Schools Eleven elementary schools	This project was designed to bring a literature-based program to elementary transition classrooms, to promote an engaging curriclum for second language learners, and to better document their learning over time. Some teachers have designed portfolios for use in the classroom and for demonstrating student achievement to parents.
	New Standards Project San Diego City Schools Sixteen elementary, middle and high schools	The New Standards Project was designed to produce a national system of standards and assessment developed by teachers and to improve instruction, build capacity, and improve access to education on the local, state, and eventually national level. San Diego City Schools is one of six districts in addition to seventeen states working with the New Standards Project. The project is beginning with tasks in literacy and numeracy at the fourth and eigth grade levels. It is funded by the Pew Charitable Trust and the MacArthur Foundation.
	UCLA/CRESST Pilot in Social Studies San Diego City Schools Three high schools	This project was designed to increase the use of primary source materials in classrooms and involve teachers in assessment development and research. Working in collaboration with UCLA/CRESST, participating teachers are piloting, adapting, and scoring tasks in U.S. History at the eleventh grade level.
	California Learning Record (CLR) San Diego City Schools Lakeside Union Elementary	The CLR is a multi-faceted student assessment instrument using teacher observations, student/parent interviews, and performances in language arts and mathematics. Teachers use the CLR to plan instruction and communicate student progress to parents. The CLR is also being studied as a possible alternative for Chapter 1 assessment.
Project in Both Consortia	**Middle School Portfolio Development** San Francisco Unified and San Diego City Schools Six middle schools	Working with Dr. Dennie Wolf at PACE, Harvard Graduate School of Education, teachers are developing literary and investigative portfolios, performance tasks, and interactive interviews. The portfolios are used to communicate growth over time to parents and students, to record growth for teachers, and in some cases, for program evaluation. The project is funded by the Rockefeller Foundation.

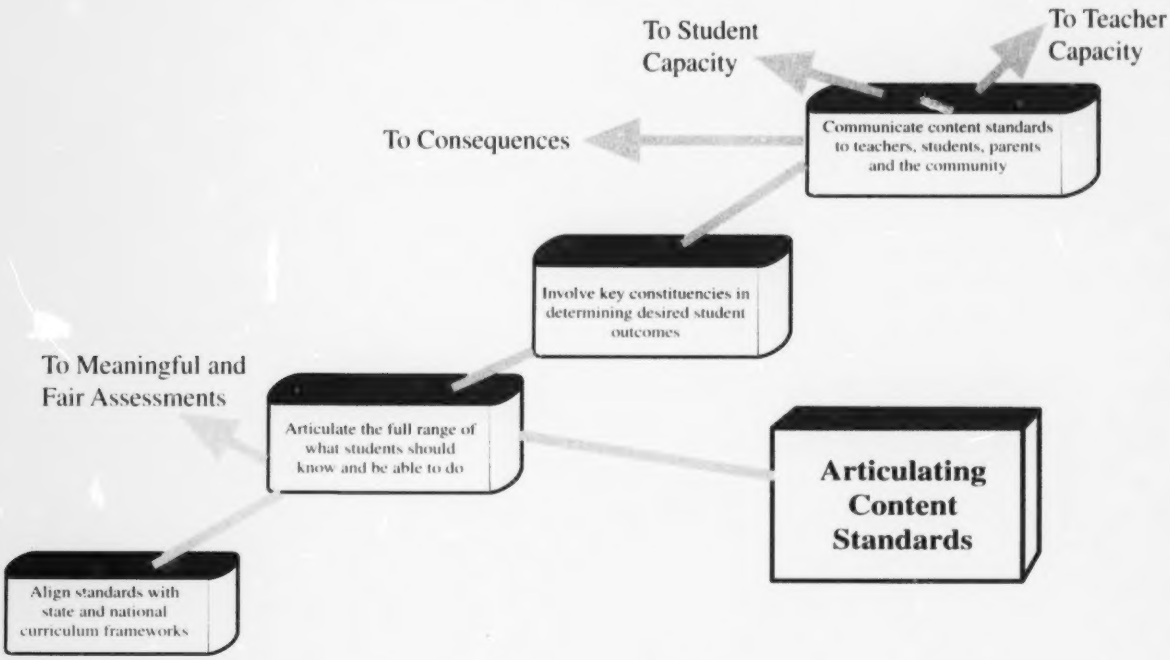
Greater San Diego Area Consortium - San Diego County Pilots

Elementary Portfolio Assessment Project La Mesa/ Spring Valley School District Fifteen elementary and middle schools	Teachers in grades two through eight are using portfolios to assess student progress in English language arts. This project was originally designed for use in GATE cluster classrooms to promote student reflection and to communicate student achievement to students, parents, and teachers. Teachers use portfolios to build evidence for student grades.
End -of-Course Assessment Project Grossmont Union High School District Ten high schools.	Teachers have developed performance-based items for inclusion on district-wide end-of-course assessments in mathematics. These items were designed to expose teachers and students to performance tasks and as a means of program evaluation. In addition, teachers in the district have developed a portfolio to assess student achievement in a new integrated science course. The portfolio is a compilation of performance tasks completed throughout the course and students' reflection on their work.
GATE Identification Portfolio Project Oceanside Unified All schools in district	K-12 teachers are developing a portfolio to be used in the process of identifying gifted and talented students. The portfolio is used by the GATE identification committee to decide if the evidence presented supports enrollment in the GATE Program.
LEP Student Assessment Project Poway Unified Eight elementary, middle, and high schools.	K-12 teachers plan to use portfolios and teacher observations to inform decisions about designation of Limited English Proficient (LEP) students, and their readiness to work in mainstream classrooms. Teachers use portfolios as a tool for exploring issues of providing appropriate instruction to LEP students, and for describing achievement to students and parents.
Math/ Science Comprehensive Assessment Project Chula Vista Eleven elementary schools	K-6 teachers have developed performance-based assessments in mathematics and science. The assessments are used and scored by individual teachers to enhance classroom instruction.
Chapter 1 Alternative Assessment Project Lakeside Union Elementary	Teachers developed criteria and assessments for identification of Chapter 1 standards.

CHAPTER 2

A CONCEPTUAL MAP

Articulating Content Standards: *What Should Students Know and Be Able to Do?*


To Student Capacity
To Teacher Capacity
To Consequences
Communicate content standards to teachers, students, parents and the community
Involve key constituencies in determining desired student outcomes
To Meaningful and Fair Assessments
Articulate the full range of what students should know and be able to do
Articulating Content Standards
Align standards with state and national curriculum frameworks

In 1993 it would be hard to find an educator, a parent, or a politician who would question the need for schools to articulate high standards and almost equally hard to reach agreement on what those standards ought to look like. Definitions of three broad types of standards have emerged from efforts to develop national standards and performance assessment strategies: *Content standards* evolve from answers to the question, "What do we want students to know and be able to do?" and *performance standards* from the additional question, "How well should they be expected to do it?" *Delivery* or *Opportunity to Learn standards* evolve from answers to the question, "To what curriculum and instructional opportunities should all students have access in order to assure a fair assessment of the degree to which content and performance standards have been achieved?"

Strategies for defining performance standards are discussed in the section entitled "Developing Meaningful and Fair Assessments." The sections on building the capacity of students and teachers to use assessment to improve teaching and learning provide an opportunity to consider some essential delivery standards. In this section, we will explore how districts and schools determine what students should know and be able to do, including

- how to articulate the full range of desired student outcomes,
- how to operationalize student outcomes so that they can be used by teachers to plan instruction, and
- how to communicate content standards to parents and the community.

The move to develop performance assessment is, in part, motivated by recognition that the content of traditional assessments addresses a narrow range of low-level thinking skills and knowledge. A recent study funded by the National Science Foundation (NSF) found that "only a handful of the questions on major commercially available tests in mathematics and science, as well as on tests included in textbooks,

A WORK IN PROGRESS

Figure 1
General Learner Outcomes
Berryessa Union School District

Students Will Be:

Effective Communicators
Students Will:
- Comunicate effectively through a variety of media.
- Effectively use a variety of information-gathering techniques and information resources.
- Effectively interpret and synthesize information.
- Be effective communicators.

Effective Problem-Solvers
Students Will:
- Recognize and analyze problems.
- Consider strategies for problem solutions.
- Propose and justify problem solutions.

Group Contributors
Students Will:
- Work collaboratively with others.
- Demonstrate effective interpersonal skills.
- Self-monitor their behavior in groups.

Responsible Citizens
Students Will:
- Demonstrate respect for others.
- Demonstrate responsibility.
- Engage in activities which address local and global issues.
- Contribute to solving environmental problems.

Life Planners
Students Will:
- Demonstrate respect for themselves.
- Create and use an educational plan to move purposefully toward career goals.
- Develop and utilize strategies that promote physical and mental health.
- Use the tools of technology.
- Establish personal priorities and achievable goals; monitor and evaluate own progress.

Developed by District Curriculum Committee with assistance from McREL and Aurora Public Schools, Colorado

measure the kinds of conceptual knowledge and problem-solving abilities which should be integral to instruction in those fields." (Madaus, 1992) Similar criticisms have been leveled at traditional assessments in other content areas.

The NSF study also found that the traditional assessments had a profound effect on what teachers taught and students learned in the classroom. The ever-increasing consequences of low test scores encouraged teachers to emphasize the low level skills required to complete multiple-choice items, taking time and energy from any efforts to facilitate higher-order skills.

If students have opportunities to learn what will be assessed, we should assess all that we consider important for students to know. An assessment system which promotes meaningful student achievement must be founded on content standards which map the full range of knowledge, skills, and learning habits we seek as outcomes for students. A major challenge in developing performance assessment is defining a comprehensive list of what we want students to know and be able to do, and then translating these desired outcomes into content standards which guide instructional planning. Described below are some of the ways CAC pilot projects and their districts have taken up the challenge of mapping these student outcomes and content standards.

District-level Efforts to Define Student Outcomes

Professional curriculum organizations are working to establish national standards in their disciplines. In California, the state subject area frameworks serve as the genesis for content standards. The articulation of those standards is a high priority of the new statewide assessment system, but is as yet an unrealized goal. Meanwhile, many local districts, often as an element of strategic planning, or as the foundation of an assessment reform effort, have begun collaborative community-wide efforts to establish a comprehensive answer to the question, "What do we want our students to know and be able to do?" Examples from two different districts demonstrate this effort.

The Berryessa Union School District has articulated both "general learner outcomes" and "program outcomes." The general learner outcomes are grouped into categories including Students as Effective Communicators, Effective Problem-Solvers, Group Contributors, Responsible Citizens, and Life Planners. The program outcomes are organized by the academic disciplines including mathematics, history/social science, language arts, physical education, science, and visual and performing arts. Figure 1 shows an excerpt from Berryessa's Student Outcomes document.

The effort to articulate learner outcomes in the Palo Alto Unified School District resulted in lists organized into "Academic and Intellectual Competencies" and "Personal Qualities." (Figure 2) The student outcomes are supported by "K-12 Outcomes of Instruction" in traditional academic areas as well as applied academics, foreign language, and health. The outcomes of instruction detail the key concepts, principles, processes, facts, and skills students are expected to know in each discipline. Both the Berryessa and Palo Alto school districts plan to use the Student Outcome documents to guide development of the districts' assessment systems.

These two lists reveal many differences in how the districts think about learner outcomes, yet both are similar in the emphasis they place on the ability to "demonstrate" or "use" academic knowledge and skills, and in the attention given to skills and abilities which cross disciplines and apply to employability and social and personal responsibilities. In articulating and building consensus around a complex range of student outcomes, these districts are laying the

A WORK IN PROGRESS

Figure 2
Student Outcomes
Palo Alto Unified School District

INTRODUCTION

We want our young people, upon graduation from the Palo Alto Unified School District, to be: **prepared for meaningful participation in an interdependent global community.**

To do so, students must develop the knowledge, skills, and values to be:
lifelong learners, responsible citizens, employable, productive workers.

OUTCOMES

As a result of their experience in the Palo Alto Schools, students will be able to demonstrate, each to the best of his or her ability, the following:

ACADEMIC AND INTELLECTUAL COMPETENCIES

- Communicates effectively through listening, speaking, and writing.
- Reads with understanding
- Demonstrates knowledge of key concepts, principles, processes, facts, and skills in each discipline: These disciplines include language arts, history-social science, mathematics, science, physical education, visual and performing arts, foreign language, applied academics, health, and practical living skills.
- Demonstrates ability to integrate knowledge
- Thinks creatively
- Thinks critically
- Utilizes technologies efficiently and creatively
- Uses research skills effectively
- Works effectively in groups
- Learns independently
- Uses effective decision-making processes
- Solves problems effectively
- Demonstrates ability to self-reflect and self-evaluate

PERSONAL QUALITIES

- Demonstrates effective interpersonal skills
- Exhibits concern and respect for others
- Appreciates the strengths and contributions of diverse cultures, groups, and individuals to our society and world
- Demonstrates social and environmental responsibility
- Demonstrates aesthetic awareness and appreciation
- Demonstrates emotional fitness
- Demonstrates a healthy lifestyle
- Exhibits personal integrity
- Exhibits positive self-esteem
- Exhibits an understanding that there are core values and beliefs which unify our community and our nation
- Exhibits a belief that learning is a joyful and rewarding life-long activity.

foundation for an assessment system which reveals the limitations of traditional standardized assessments. Although those tests may provide some data on a very limited number of student outcomes, it is also clear the communities value and want assessment data on many important outcomes which traditional assessments ignore.

Operationalizing Student Outcomes — Articulating Content Standards

Student outcome statements, although comprehensive, will not guarantee improved student performance until teachers themselves have articulated the way those standards will be applied to their own instructional program. Content standards become valuable instructional tools when teachers articulate the specific knowledge, skills, and learning habits they must teach in order for students to achieve the stated outcomes.

Several CAC pilot projects offer examples of ways in which student outcomes are operationalized by articulating content standards.

Working from Curriculum Frameworks - The Values Exercise

California's curriculum frameworks are seen as models by other states and across the nation. Although they do not specify comprehensive content standards, their vision gives strong direction to those attempting to articulate what students should know and be able to do. The frameworks have in common a foundation in recent cognitive research which describes learning as the act of constructing meaning from new information through prior knowledge. Each calls on educators to teach facts and skills as tools for meaning-making, and not as ends in themselves. The frameworks challenge educators to determine discipline-specific core curricula while encouraging the exploration of cross-disciplinary themes. Each is grounded in a commitment that all students must have access to and succeed in a demanding intellectual environment.

Several CAC pilot projects have employed "The Values Exercise," an activity first used in the ACE Consortium K-2 Assessment Project to determine what students should know and be able to do. The activity has been found useful in helping teachers translate concepts presented in the state curriculum frameworks into a comprehensive list of valued essential questions and content standards which can be used to plan instruction and assessment in their own classrooms and schools.

Working in grade-level clusters, course-alike, or subject area groups, teachers answer the question, "What knowledge, skills and habits of mind do you value in students as writers . . . as students of U.S. History, etc.?" They answer by brainstorming to produce a list of questions which begin, "Can the student....?" or "Does the student...?" Each question suggests valued content.

When they have exhausted the scope of knowledge, skills and learning habits they consider important to their field or course, teachers are then asked to group the questions into categories which can be defined by a broader "umbrella question." The task of grouping the umbrella questions is repeated until participants arrive at questions which describe the essential aspects or core learnings of the curriculum. The umbrella questions describe the scope of that curriculum, and the questions generated in the initial brainstorm serve as guides to behaviors which indicate progress toward achievement of skills asked about in the umbrella and core questions. The resulting document, arranged as core questions, followed by "umbrellas" and the related questions generated in the initial brainstorm, becomes the content standards on which assessments or a system of assessments can be based. (An example of the core and umbrella questions for a K-2 language arts program as determined by the ACE K-2 project can be seen in Figure 3.) Participants then move on to develop meaningful tasks to assess these established content standards.

In the first stage of the exercise, teachers often brainstorm a diverse range of questions which include important, but low-level skills (i.e., Does the student capitalize proper nouns?) along with more abstract aspects of the field (i.e., Does the student write with a sense of audience?). At this stage of the exercise, participants are not asked to consider the level of importance or emphasis they attribute to any given question, nor required to agree on which should be included. All questions considered valuable by any participant are recorded.

When participants arrange these original questions into categories and begin to negotiate the phrasing of "umbrella questions," they begin to negotiate the emphasis given to various aspects of the content, and to construct a meaningful collective understanding of concepts from the state curriculum frameworks.

A Work in Progress

Figure 3
Essential Questions — ACE K-2 Alternative Assessment Project

As a reader	As a writer	As a listener/speaker
Does the child approach language tasks with a positive attitude and a willingness to take risks?		
Does the child enjoy reading and see herself as a reader?	Does the child choose to communicate through writing? Does the child take initiative as a writer and see herself as a writer/author?	Does the child value what she and others have to say? Does the child appreciate the aesthetics of language?
Does the child use strategies to make herself understood and to derive meaning from what others have to say?		
Does the child use strategies to derive meaning from print? Does the child recognize patterns in literature/text and use them? Does the child consult appropriate resources?	Does the child see writing as a process? Does the child have strategies to improve her writing? Does the child use writing to explore or organize ideas she hears or reads?	Does the child use strategies for speaking and listening effectively?
Does the child have increasing control over the conventions of English?		
Does the child understand the conventions of print?	Does the child use the conventions of grammar, spelling, legibility, and punctuation correctly?	Does the child use the conventions of spoken English appropriately?
Does the child have a sense of audience and approach language tasks as a problem solver?		
When reading to an audience, does the child vary her intonation and pace to make the text meaningful and interesting?	Does the child use strategies to help the reader understand or maintain interest in her writing? Does the child have strategies to improve her writing?	Does the child use strategies to help the audience understand or maintain interest in what she has to say?
Does the child use language effectively to express original thoughts and for a variety of other purposes?		
Does the child choose from a variety of genres? Does the child use reading for a variety of purposes? Does the child construct personal meaning from what she reads?	Does the child use writing effectively for a variety of purposes? Does the child express original thoughts in writing?	Is the child an effective listener? Does the child speak effectively in a variety of settings for a variety of purposes? Does the child speak using original thought?

Pushed to be comprehensive about what they want students to know and be able to do, teachers list not only traditional skills but learning habits or "habits of mind" essential to the field. For example, teachers responsible for mapping content standards in writing will usually arrive at essential questions which ask about the student's ability to take risks, to persist through difficult assignments, or to use writing as a tool for exploring thinking, not just reporting on it. When stated as valued content, these habits of mind become as much a part of what it means to write well as does the ability to use punctuation correctly.

Several CAC project schools have used the Values Exercise to generate a comprehensive description of the desired school-wide curriculum in a given content

BEST COPY AVAILABLE

area. In order to do so, they have included teachers and other district staff not originally involved in the alternative assessment project. Some schools report plans to use the exercise with school site councils and other parent groups in an attempt to broaden the base of support for curriculum reforms.

Working From A Variety of Standards Documents

The Mount Diablo Unified School District called on existing standards documents to help articulate the foundation of their performance assessment efforts. Project leaders compared the knowledge and skills assessed on tests already in use in the district, with the curriculum standards developed by the National Council of Teachers of Mathematics in planning to enhance the district-wide assessment of mathematics in grades 3 and 6. Teacher participants noted that both the NCTM standards and the California Mathematics Framework called for students to demonstrate knowledge of "Attributes and Classifications," "Dealing with Data," and "Locating, Mapping and Planning Space," yet none of the assessments currently in use in the district addressed these standards. The resulting Curriculum-Based Alternative Assessment for Mathematics was designed to include open-ended performance items to assess skills in those areas.

The California Learning Record (CLR) is an assessment and instructional tool modeled on the British Primary Language Record. Teachers and students keep a portfolio of student work accompanied by notes which record teacher observations. The teacher observation document itself reflects content standards by defining the range of desired skills and habits for which teachers are charged with finding evidence.

Starting Over — Articulating Content Standards through Course Development

Teachers in the Coordinated Science Program at Monte Vista High School in the Grossmont Union High School District were challenged to articulate content standards in the process of developing a two-year integrated science course designed to cover the same knowledge and concepts as existing physics, biology, chemistry, and physical science classes. Using the previously established end-of-course exams and achievement tests in each content area as anchors, teachers looked for common themes which were woven through the traditional disciplines and developed integrated thematic instructional units and performance tasks congruent with each. In the act of integrating multiple, previously established courses, teachers are forced to consider what learnings are essential and what the scope of the new course should be. As their work continues, Monte Vista teachers have developed a research plan to determine whether student performance in the coordinated science course equates with the performance of students in the traditional courses.

A similar effort to establish an eighth grade humanities course which integrates history/social science and language arts at Muirlands Middle School in San Diego City Schools led to content standards expressed in terms of core and essential questions. Each essential question is explored through a study of certain periods in American history and readings from a congruent core literature list. Students are challenged to complete interdisciplinary projects related to each question. The content standards for the interdisciplinary course require not simply the memorization and recall of facts or rules, but the use of knowledge applied to complex performance tasks. An outline of the course content and essential questions are seen in Figure 4.

In the examples above we see several different ways of going about the process of articulating student outcomes and content standards. District-level efforts are designed to assure commitment to outcomes which include all academic disciplines, employability skills, and personal and social responsibility. Within specific courses or content areas, the articulation of standards assists teachers to operationalize those outcomes for instructional planning. All these efforts recognize that performance assessment development is not solely an effort to create new assessment tools, but to assess a wider range of outcomes than traditional methods can address.

Communicating Content Standards to Parents and the Community

Maintaining support for continued performance assessment development work will require that the breadth of knowledge, skills, and learning habits to which our instructional programs aspire is made clear to the public and policy makers. When parents understand the full scope of what students are expected to know and be able to do, they are not only more likely to support reforms aimed at preparing students appropriately, but they will be better equipped to support work toward those standards at home.

A Work in Progress

Figure 4
Grade 8 Humanities Course Outline
Muirlands Middle School
San Diego City Schools

Year Question: Who are we Americans?

Essential Questions:
I. How does the past influence who I am? 1500-1800
  - Constitution
  - Exploration
  - Colonial Including American Revolution

My Brother Sam is Dead, Sarah Bishop, April Morning, The Crucible

II. What happens when cultures collide? 1800-1865
  - Westward Expansion
  - U.S. / Mexico
  - California / San Diego (Gold Rush)
  - Western Day

The Ox-Bow Incident, Shane, When the Legends Die, The Light in the Forest, The Way West

III. How does conflict affect us? 1800-1865
  - Slavery (historical antecedents and 19th century realities)
  - Regional Differences (social, economic, political)
  - Civil War

Across Five Aprils, The Killer Angels, Huckleberry Finn, Frederick Douglass

IV. How did the "American Dream" arise? 1865-1900
  - Reconstruction
  - Inventions and Technology
  - Urban / Rural Issues
  - Immigrants

Freedom Road, Call of the Wild, I Heard the Owl Call My Name, A Night to Remember

V. Can one person make a difference? 1900-Present
  - Civil Rights
  - Discrimination
  - Judicial Power
  - Legislative Power

To Kill a Mockingbird, Twelve Angry Men, Gideon's Trumpet

Teachers within several CAC pilot projects make explicit efforts to communicate content standards to parents as part of the assessment process. An example is provided by one fifth grade teacher in the writing portfolio project at Stevens Creek School in the Cupertino Union School District. After students complete an initial draft of a "high fantasy" narrative, parents receive a letter (Figure 5) which guides their responses to the child's writing. Responding to the teacher's letter requires parents to assist their children with articulating their ideas, and allows them to better understand the actual expectations for student performance.

The California Learning Record includes, as a part of the assessment portfolio, an interview with parents about the child's literacy experiences at home and at school, and an opportunity for parents to comment on the Record completed by the teacher. By involving parents in these ways, teachers using the Record are better able to communicate their expectations for literacy development.

Challenges, Cautions, Caveats

The Challenge of Being Explicit about Content Standards

It is not uncommon for an assessment development effort to begin as a "portfolio project" or a "performance assessment project," establishing the assessment technology before any other decisions are made. Participants may spend hours debating what should go in a portfolio, or the characteristics of a good task, without ever having established what they hoped to assess in the first place. It is often assumed that everyone knows, or shares, a general understanding of a stated student outcome or of what is important for students to know and be able to do within a given content area.

Observations across the CAC pilot projects suggest that this assumption of shared outcomes is seldom true. Perhaps because most teachers have so long been removed from decision-making about curriculum content and have rarely been asked to articulate anything other than the most easily measured outcomes, the work of articulating comprehensive standards can be extremely challenging. The exercise itself, which takes time, expertise in content and learning theory, and a willingness to debate and negotiate, challenges the cultural and structural norms of most schools. In most schools teachers are isolated from one another, have little access to state-of-the-art knowledge and

A Work in Progress

Figure 5
Letter to Parents — Stevens Creek School

Dear Parents,

We have been reading fantasy stories and are now writing a fantasy of our own. The genre of "high fantasy" is one that the children really enjoy writing! We have focused on what makes a good fantasy and have used the literature to find out. We have done brainstorming in order to define characters and setting, so these are the focus of the assignment. Below is a checklist of outcomes we have defined which make up a good fantasy.

Please have your child read you their "sloppy copy" of their fantasy. Fill out the checklist below and ask your children questions about their story, also tell them what really worked! The students have great imaginative ideas. It is important that they talk about their stories in order for them to be able to clarify what it is they want to get across.

Thanks for all the support, time, and enthusiasm you give your child. I couldn't do it without you!

SETTING:

- Included magical features:
  - list the most imaginative, intriguing magical features that were a part of the author's setting
  - list some of the vocabulary that the author used which helped you see this magical place

- Included an adventure through different places, lands, or experiences
  - list the different places that the characters went through on their adventure and tell how you would feel if you were in those places
  - list some of the vocabulary that the author used which helped you understand what kind of a place it was for the characters

CHARACTER:

- Characters are clearly defined
  - list the characters in the story and write an adjective after each which describes him/her/it; also write whether the character is a protagonist or antagonist

Character	Adjective	Protagonist?

- The author uses an omniscient point of view as well as descriptive language to show:
  - the characters' thoughts and reactions about the setting - list an example of the author doing so, list any suggestions
  - the characters' thoughts and reactions about other characters - list an example of the author doing so, list any suggestions
  - the characters' thoughts and reactions about the sequence of events - list an example of the author doing so, list any suggestions.

- The author was most effective in which of the following areas? Tell why.
  - creating for the reader a magical land which has imaginative characters, magical features, and a variety of places and experiences
  - creating a set of characters in which the reader could understand exactly who they were because:
  - physical features were described well
  - their thoughts and reactions to other characters and the setting were described well.
- Please draw a picture of the protagonist in the story and an antagonist in the story.

Developed by Caroline Embrey, Teacher, Cupertino Union School District

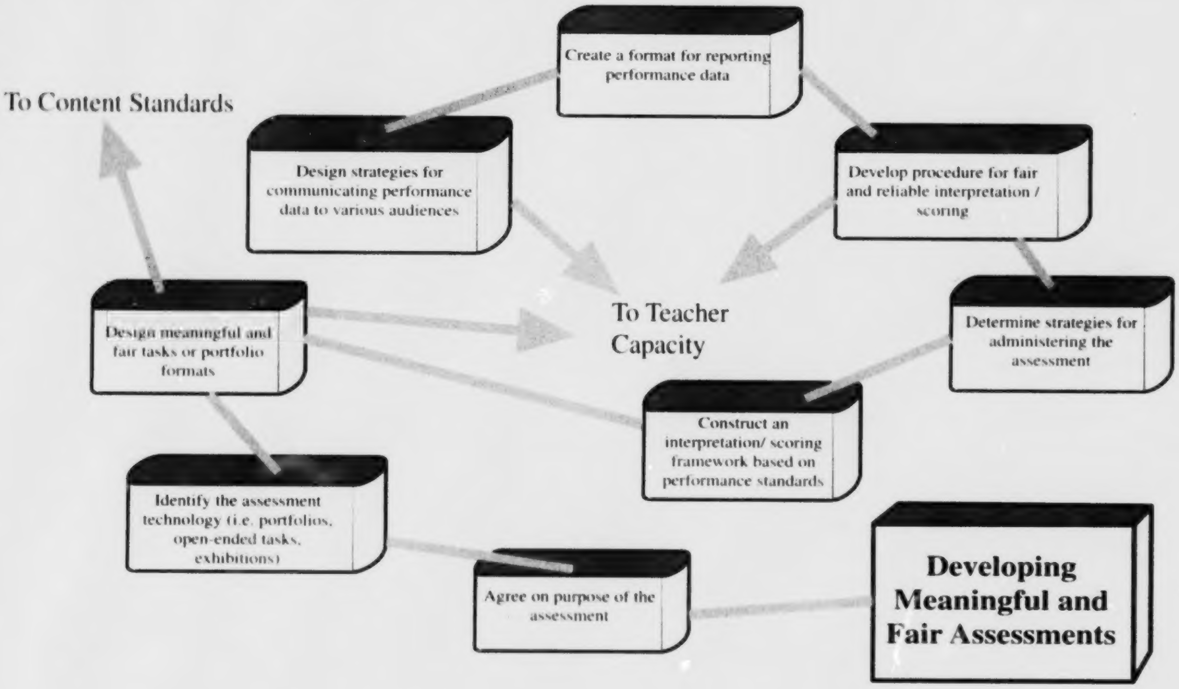
resources, and seldom risk substantive disagreements with professional colleagues. As we shall see in Chapter 4, participation in performance assessment development efforts can be the catalyst for these critical conversations, and for challenging existing structures and cultural norms.

Moving on to Performance Standards — Establishing How Good is Good Enough?

Establishing what students should know and be able to do is an essential step in the development of performance assessment, but it is also important to define how well students are expected to perform. There is little precedent in schools for encouraging a consensus around the level of student performance we should expect. Grading has always been a uniquely private act. Individual teachers determine the criteria for good performance, and they may or may not communicate those criteria to students, let alone to other teachers. It is commonly assumed that work produced in one class may result in a different grade in a similar course taught by a different teacher, and that a given grade average in one school may describe a different level of achievement in another.

One of the most challenging aspects of developing alternative assessments is to agree upon and explicitly state what it means to do well. Schools must negotiate and build a consensus around the criteria which will be used to judge student performance. Some of the ways CAC pilot projects have gone about this work are described later in this chapter when scoring frameworks and processes are discussed.

DEVELOPING MEANINGFUL AND FAIR ASSESSMENTS:
HOW DO WE DESIGN ASSESSMENTS WHICH ENCOURAGE THE BEST FROM ALL STUDENTS?


Create a format for reporting performance data
To Content Standards
Design strategies for communicating performance data to various audiences
Develop procedure for fair and reliable interpretation / scoring
To Teacher Capacity
Design meaningful and fair tasks or portfolio formats
Determine strategies for administering the assessment
Construct an interpretation/ scoring framework based on performance standards
Identify the assessment technology (i.e. portfolios, open-ended tasks, exhibitions)
Agree on purpose of the assessment
Developing Meaningful and Fair Assessments

It should come as no surprise that participants in assessment development efforts initiated and / or sustained primarily by classroom teachers name as their primary audience "the classroom teacher" and their primary purpose "the improvement of teaching and learning." Faculty rooms have long been filled with talk of the paltry impact of traditional large scale assessments on instructional practice. Teachers speak of the mismatch between what they teach and what is tested, the limitations of "one-point-in-time" analysis of student performance, the long delays in receiving feedback on results, and the inaccessible, often obscure, ways in which norm-referenced data are reported.

In the last ten years, these informal conversations have been amplified by a chorus of researchers and policy makers. Backed by significant research, these voices argue that traditional assessments not only add little to, but actually detract from, student learning.

Various commentators cite evidence that traditional assessments

- instruct teachers to narrow content to a limited range of low-level skills,
- suggest that questions can be answered in only one way,
- emphasize selection rather than creation of a response,
- are biased against various groups or types of students,
- carry consequences that inhibit motivation and opportunities to learn, and
- take valuable time away from instruction.

(Wigdor and Gardner, 1982; Madaus, 1992; Resnick, 1989)

Teachers participating in CAC efforts to develop assessment alternatives speak of their motivation to do so in terms of creating assessments which

- provide timely and usable feedback to students and teachers,
- seek to capture evidence not only of skills and knowledge, but the ability to consistently apply those to real world problems,
- provide incentives to teach and learn what's really important, not just what is easily measured,
- are fair for all students, and
- carry consequences that increase motivation and opportunities to learn.

In short, their efforts are sustained by the belief that although performance assessments may take more time, they will be "instructionally sound" and can be used in an on-going way to effectively plan meaningful instruction for the class or to design interventions for individual students. In addition, recognizing that assessment also serves an influential gatekeeping function, many sponsors of and participants in performance assessment development efforts also hope to devise tools for generating reliable class, school, or district program evaluation data or individual student scores. The hope is that a single assessment might be instructionally sound, reliable, and efficient enough to be used in high stakes program evaluations or decisions affecting the participation or advancement of individuals in various aspects of the school program.

Technologies of Instructionally Sound Assessment

The most common formats for traditional assessment use technologies such as multiple-choice or true/false questions or short answer/fill-ins. Such assessments are administered as the culminating activity of an instructional unit or block of school time. Those developing alternatives seek technologies which allow students to demonstrate what they know and can do using strategies which are both learning and assessment experiences. The call for assessment tools which positively affect teaching and learning has given rise to the following categories of alternative technologies or tools: portfolios, on-demand tasks, curriculum-embedded assessment, projects, and exhibitions. Each is described below:

Portfolios - The term portfolio usually refers to a purposeful collection of student work selected over the length of a course or school year. Portfolios might contain work required and selected by the teacher and/or students. Work collected may be unique to the individual student or may represent similar tasks assigned to all students in a class or across classrooms.

Typical formats include "best piece" or "showcase" portfolios which require students to select what they consider to be their best work for review and evaluation. "Menu-driven" portfolios require students to compile entries covering a variety of types of assignments (i.e., six open-ended questions, two lengthy investigations, one essay of self-evaluation) and/or representation from various units of study. Portfolios may also be organized by a set of concepts or valued learnings for which students are responsible for finding evidence in their work (i.e., the ability to solve problems cooperatively, the ability to develop two sides of an argument).

Other features of a portfolio often include

- reflections or student self-evaluations,
- feedback from parents or other "third party" observers of student work,
- opportunities to demonstrate knowledge or abilities through media other than writing (e.g., videos), and
- anecdotal records of teacher observations.

On-demand tasks - This term refers to performance tasks which are of short duration, completed within one, or only a few, class periods. Students are required to respond to a common task within a given time period. All students are provided the same directions, resources, and time to complete their work.

Curriculum-embedded assessments, projects, and exhibitions - These often-interchangeable terms refer to assessment tools which mirror instructional practice and usually take place within the classroom over the course of several class periods or even weeks. Curriculum-embedded assessments invite different students to respond to a single prompt in a variety of ways. They may also require collaborative group work.

Plans for the new California Learning Assessment System call for the use of all three of these types of assessments. The CAC's first year pilot projects provide examples of all these assessment technologies. Although evolving work is difficult to categorize, the assessment technologies used by pilot projects are described beginning on the next page.

What We're Learning About the Development Process

Each of the CAC's projects followed its own pattern of development at its own pace. Starting points were as different as the assessments which evolved. Developers of some projects began by deciding to employ a given technology and then developed items or formats they thought appropriate to the grade levels or subject areas on which they would focus. Others began by determining what they wanted students to know and be able to do and then determined the technology they believed could best be used to assess those things. Some project developers thought first about the purpose and audience for new assessments and then developed a technology, scoring process, and reporting procedure aligned to that purpose. Others adopted or adapted assessments or parts of assessments developed elsewhere and focused on implementing them in the local setting.

Although no absolutely "right way" to proceed through the development process has emerged from our work, collectively, the CAC's first year pilot projects have helped us identify essential elements of the development of meaningful and fair assessments, including

- determining the purpose for assessment,
- exploring the relationship between technology and purpose,
- designing tasks, formats and standards,
- considering strategies for administration,
- designing interpretation or scoring frameworks,
- developing performance standards,
- structuring reliable scoring procedures, and
- designing a format for reporting data.

The experience of CAC projects suggests that the choices made at these critical junctures can affect the degree to which the assessment positively influences teaching and learning.

Determining Purpose

Few of the declared purposes of CAC pilot projects are distinct, nor are they necessarily static. Although it is accurate to say that nearly all CAC pilot projects strive to provide assessment data which are accessible and useful to classroom teachers for the primary purpose of improving teaching and learning, it is true that many aspire, in addition, to create reliable gatekeeping or accountability tools. Within any one project there may be individuals who perceive their purpose to be all of these.

The Mt. Diablo Curriculum-Based Math Assessment provides an interesting example of the multiple and evolving purposes for new assessments. The project was initiated as an effort to help teachers and students prepare for CAP. Designed by the staff of the district's Research and Evaluation Office with assistance from curriculum coordinators in mathematics, the project was intended to provide teachers and students experience with open-ended problems and the district with data for school accountability reports. Initially, the district planned to provide each school with a detailed public report of performance results. Reports from the spring administration of the performance-based matrix-sample test would be available to schools the next fall.

Recently, the project has placed greater emphasis on affecting instructional practice. In the 1992-93 school year, the test was given in early March to provide principals with class and student scores to be used in instructional or program planning. Other changes in plans for the 1992-93 administration included holding orientation sessions for all teachers on strategies for administering the test. Teachers were encouraged to give students similar resources and assistance to ensure that the test was an effective learning experience.

Other projects provide examples of the ways assessments, originally developed purely to influence instruction, have begun to be considered for use for accountability. The Primary Language Record was developed in England for the purposes of monitoring and reporting growth in learning. Adapted as the California Learning Record, the instrument is now under study as a means of monitoring the success of Chapter 1 programs by documenting student progress.

CAC Pilots: Tasks, Technology, Scoring

Portfolio Description	Scoring Procedure / Purpose and Use:
Portfolios with a common template: **These portfolio assessments include a common template or table of contents, and at least some common tasks or specifications for common tasks.**	
ACE K-2 (Selected Bay Area Schools)	
Portfolio template includes common project-like tasks, and observation and interpretation forms.	Individual teachers interpret performance against commonly developed performance standards. Interpretation results in an instructional plan for the student and teacher.
O'Farrell Community School, PACE Project, (San Diego)	
Portfolio contents require students to select evidence of achievement of agreed upon content standards. All students respond to common project protocols. Portfolio entries require self-evaluation and "third party" feedback.	Individual teachers score student work, some using task specific rubrics. Portfolio is presented to panel of reviewers where achievement of the "O'Farrell Standard" is determined.
Ben Franklin Middle School, PACE Project, (San Francisco)	
All students keep classroom work folders and select work for school level portfolios compiled twice a year as evidence of what they've learned. Tasks are not standardized, although some projects are required of all students taking a certain course. All portfolio entries require self-evaluation and many request parent feedback.	All tasks are graded by individual teachers, some using task-specific rubrics. A student presentation of the portfolio before a panel of reviewers has been field tested, but not yet implemented. Faculty use contents of student portfolios for informal program reviews.
Grossmont High School District Coordinated Science Course	
Portfolio contents require students to select evidence of achievement of agreed upon content standards. All portfolio entries require student self-evaluation. Many tasks are common across coordinated science classrooms.	Students score own portfolios using a common rubric. Teachers re-score. Portfolio score is used as a major part of the student's grade.
Classroom Portfolios: **The following portfolios provide templates and content standards but are used to compile work unique to each classroom. The portfolio is used to capture evidence of what is already happening in the classroom and does not require common tasks.**	
La Mesa -Spring Valley School District	
The portfolio template suggests grade-level appropriate genres of student writing and content standards for those genres.	Teachers individually grade student work. At this writing there is no common scoring strategy for portfolio as a whole.
California Learning Record (San Diego City, Lakeside)	
Adapted from the Primary Language Record, the CLR guides teachers and students to collect student work and records of observations which serve as evidence of achievement of content standards. Some evidence comes from parent interviews or reflections. Older students are responsible for selecting pieces from their own classwork. Student tasks are not standardized.	CLR includes reading, writing scales which are used along with observation forms by individual teachers to report student achievement .

Portfolio Description	Scoring Procedure / Purpose and Use:
The following portfolios provide templates and content standards but are used to compile work unique to each classroom. The portfolio is used to capture evidence of what is already happening in the classroom and does not require common tasks. (Cont.)	
Stevens Creek School (Cupertino)	
Students keep both working and showcase portfolios. Selections for the showcase portfolio are made by students with feedback from parents and teachers. No standardized or common tasks are required.	Multidimensional-rubric for narrative writing is used to interpret individual assignments. Teachers individually grade other work. No common strategy has yet been developed for scoring portfolio as a whole.
Oceanside GATE Portfolio	
Teachers compile portfolios to provide evidence of agreed upon content standards or characteristics of exceptionality. No common tasks required.	Teacher compiles portfolio to submit to site level identification team. Team reviews portfolio and sends recommendation on to Gifted Program Administration for final decisions about identification.
These portfolio assessments had no commonly required contents or tasks and were used primarily for program planning and professional development.	
Poway Limited English Proficiency Portfolio	
Participating teachers each have unique portfolio templates. Successful tasks were shared among teachers on the development team, but not required commonly of students.	There is no common strategy for interpreting or scoring portfolios at the present time. Teachers use examination of student work to determine good instructional strategies for English learners and as a vehicle for discussing process for reclassification.
Newcomer High School (HERALD, San Francisco)	
Participating teachers each had unique portfolio templates. Teachers collectively examined common student portfolios for the purpose of program evaluation and instructional planning.	No common scoring strategy for portfolios has been used. Teachers use examination of student work to determine good instructional strategies for English learners and to assess the level opportunity to practice writing in the content areas.
These assessments made use of open-ended on-demand tasks which were designed to be given by many teachers across many classrooms.	
Redwood City Performance-Based Assessments	
Students will complete a matrix sample assessment of language arts and math tasks during the PBA assessment window each spring.	Teachers will score items in double-blind scoring sessions using task-specific rubrics. Schools will receive school-level scores. At the time of this writing it has yet to be determined how student performance data from several different items will be compiled.
Mt. Diablo Curriculum-Based Mathematics Assessment	
Students take three open-ended math tasks during a testing period in the early spring of the school year.	Volunteer teachers score items in a double-blind holistic scoring process. Teacher and school level scores are reported to the principal. Aggregate district scores are available to all. For the first time in the 92-93 school year, individual student scores will be available for teachers and students.

Portfolio Description	Scoring Procedure / Purpose and Use:
These assessments made use of open-ended on-demand tasks which were designed to be given by many teachers across many classrooms. (Cont.)	
San Lorenzo Valley Curriculum-based Assessment in Special Education	
Special Education students complete a series of reading, writing and mathematics tasks in the fall and spring of each school year.	Teachers score some tasks individually and others in double-blind holistic scoring sessions. Individual student scores are reported at IEP'S. Aggregated data are presented as part of annual Special Education Program Evaluation.
San Lorenzo Valley K-2 Assessment	
Students complete a series of reading, writing and mathematics tasks in the fall and spring of each school year.	Teachers score some tasks individually and others in double-blind holistic scoring sessions. Teachers may use individual student performance data, but only aggregated classroom level data are reported as part of the district.
New Standards Project (San Diego City)	
Fourth and eighth grade students field-tested one element of the envisioned assessment system by taking individual math and language arts items.	Pilot districts and states conduct local scoring sessions. Data are submitted to NSP for project development purposes.
CRESST-U.S. History (San Diego City)	
U.S. History teachers piloted use of written performance tasks developed jointly with CRESST which took from one to three class periods.	Teachers score items in CRESST-sponsored scoring sessions. Performance data are used for research and development only.
These assessments employed curriculum embedded, on-demand tasks or portfolios for classroom use only.	
Elementary Science & Language Arts Assessment Development (San Diego City)	
Participating teachers developed performance tasks for use in their own science or language arts programs.	Individual teachers have developed rubrics for use in scoring some tasks for their own classroom use.
Chula Vista	
A district development committee produced common on-demand, curriculum embedded math and science tasks which are distributed for use by all interested teachers.	The district development committee adapted a generic rubric to be used by individual teachers along with task-specific content criteria to score items. Some schools have experimented with collaborative scoring. No results are reported.

In other projects aimed primarily at instructional improvement, some participants reveal a hope that the assessment will also some day serve accountability. In the ACE K-2 Project, for example, some schools have used their participation in the project to argue for a reduction in traditional standardized testing in primary grades with promises of having program accountability data from the alternative assessment in the near future.

The complexity of purpose revealed in CAC projects mirrors the debates occurring throughout the assessment field. Can performance assessments be used effectively for both instructional improvement and accountability? What loss of opportunity for improved teaching and learning will result from the need to meet criteria for accountability? What loss to the reliability of accountability data will result from an increased emphasis on instructional integrity? Although our early work offers few direct answers, it does suggest that at least for now, those championing alternative assessments for the sake of accountability and those whose interests are singularly focused on instructional improvement are still willing to admit there may be something to learn from each other.

Determining the Technology Based on Purpose

Determining the purpose for which an assessment will be used can help guide decisions about what technologies to use and how they should be designed. The Oceanside GATE Portfolio project provides a good example of a project which matched its assessment technology to its purpose and audience. Traditionally, "Gifted and Talented" programs rely on standardized achievement or IQ tests to determine eligibility. When students have been identified, parents and school level personnel collaborate on a plan which guides the student's instructional program.

The GATE Task Force in Oceanside believed that academic achievement as defined by standardized test scores alone was not a valid indicator of giftedness. They sought an assessment technology which would allow students to demonstrate a variety of forms of exceptionality using a variety of media. Through the alternative technology they also hoped to have evidence that exceptional performance was not a one-time occurrence but a sustained characteristic. Finally, they looked for a technology which would give guidance to the parents and school personnel who would eventually chart the gifted student's instructional program. The GATE portfolio, structured in a way to require teachers and parents to collect and judge evidence against a variety of indicators of exceptionality, is a technology which meets all these criteria.

Other projects provide examples of the ways in which schools discover and respond to the limits of performance assessment technology when they have clearly stated their purpose. The Redwood City "All Students Will Succeed" Assessment Project is designed to provide both formative and summative assessment data. At its start, the district analyzed its Scope and Standards document to determine which knowledge and skills should be tested in the on-demand, summative Performance-Based Assessment, and which were more appropriately assessed in a formative way using technologies like portfolios. While development continues on the summative assessment, the district also supports site-level experiments with portfolios and participates in projects featuring other assessment technologies.

The Poway Unified LEP Portfolio Project and the first year work by some of the San Diego and San Francisco Middle School Portfolio Project schools demonstrate how a variety of technologies can come into play within a single effort. While focusing on portfolios, participating teachers nonetheless have put great emphasis during their early work on both curriculum-embedded and on-demand task development, considering the advantages and limitations of each within the portfolio. The operating assumption of both these projects, and many similar portfolio efforts, is that the portfolio will be only as instructionally valuable as the tasks included in it.

Developing Tasks or Portfolio Formats based on Content Standards

As described earlier in this report, an essential step in developing meaningful assessment tasks or in making good decisions about what to include in a portfolio is determining clear content standards that answer the question, "What should students know and be able to do?" Without these there are no criteria by which to judge the quality of a task or portfolio format. But assuming content standards are in place, the challenge of creating assessments to measure their achievement should not be underestimated.

Lists describing the essential characteristics of performance tasks abound, though seldom does any one task or portfolio design meet all the criteria. A list compiled from a number of resources used in CAC projects is seen in Figure 6.

Figure 6
Characteristics of Authentic Assessments

- Does the task spark student interest and motivation to do well?
- Does the task require students to construct meaning?
- Does the task require students to explore essential questions or "big ideas"?
- Does the task encourage demonstration of important "habits of mind"?
- Does the task encourage multiple modes of expression?
- Does the task encourage or require student self-evaluation?
- Is the task free from arbitrary constraints? (time, access to resources)
- Does the task measure progress over time?
- Does the task require collaboration with others?
- Is the task a representative challenge, emphasizing depth rather than breadth of response?
- Does the task explore and identify hidden strengths?
- Does the task genuinely assess learning and effort, not native talent?
- Do the standards for the task cover a wide range of knowledge, skills, and habits of mind considered important to the subject area or, if interdisciplinary, those that transcend a single discipline?
- Are the standards for good performance clear to students before they engage in the task?
- Are the standards for the task appropriately weighted?
- Are the standards for the task in harmony with shared (school, district, state, national) goals?
- Does the task match the scoring framework?
- Is the task multidimensional, allowing for a single performance to be strong in some areas and weak in others?
- Is the assessment structured to provide prompt and useful feedback to the teacher and student?
- Does the task provide built-in opportunities for students to practice, rehearse, and retake it?

(Wiggins, 1990; Mitchell, 1992)

CAC pilot projects provide endless examples of the recursive nature of the work involved in developing good performance tasks. Teacher participants draft tasks, or make collective decisions about the structure of a portfolio, try it out with students, and revise it numerous times based on the data from student responses. This process is characterized in a comment from Redwood City School District's newsletter, "The Assessment Sampler": "Developing the PBA prototypes has been an experience for us all. The writers have revised and revised, rubrics have been made and revised and revised, and scorers have scored and argued and scored again."

These action-research projects address the validity of the assessments by working to assure that the tasks engage students in meaningful work and actually require students to demonstrate the knowledge and skills they purport to assess. Examples of two drafts of a task from Mt. Diablo Unified's Curriculum-Based Math assessment demonstrate how successive revisions affect what is actually assessed.

The task was designed to assess students' ability to "deal with data":

"Five people belong to a club, and two are going to bring refreshments each week to their weekly meetings. How many weeks go by before the same two people would have to bring refreshments?

If another person joins, how many pairs can be made with six people?"

To produce a correct answer to this problem, a student could use combinatorial reasoning, trial and error, or make a lucky guess. Field tests of this item revealed that student answers seldom provided information about the students' ability to infer, recognize, and

BEST COPY AVAILABLE

display their knowledge of rules and patterns in mathematics.

In its second version, the task developers added the question, *"What is the pattern used to determine the number of pairs for any number of people?"* This addition requires students to go beyond a single correct answer and display data which reveals the process they used and how it can be used to answer a similar problem.

The required contents of portfolios also goes through revisions as developers examine how or whether that technology assesses what students should know and be able to do. Figure 7 shows the changes in the contents of the portfolio used at O'Farrell Community School in San Diego. Teachers at the school constantly revise the portfolio format as they work together to define its use and its relationship to student outcomes. The first list was made in the school's first year of operation and was based on the curriculum designed for seventh grade students. The following year the faculty redefined their interdisciplinary curriculum and therefore wanted to move away from subject-matter-specific entries to social/emotional and academic areas that transcend discipline-oriented definitions. As the O'Farrell community continues to define their teaching and learning goals, the portfolio design and required entries are likely to experience continued revision.

It is not only what students are asked to do, but how they are asked to do it that affects the validity of an assessment. The language used in task directions or prompts can greatly affect what students do and what their work actually reveals about what they know. Figure 8 shows three progressive drafts of a science task developed in Chula Vista. The purpose of the task is to assess students' knowledge of the essential elements of the cycle of evaporation and condensation.

Note that the first draft, using words such as "what will happen to you now" and "show your adventures" invites students to write about or draw imaginative occurrences which may have nothing to do with the scientific concepts the task is supposed to address. The directions ask only for a labeled drawing, suggesting no opportunity to demonstrate causal relationships or complex understanding.

By the second draft, the invitation to imagine adventures has been eliminated. This version uses language ("tell how you became") which causes the student to explain not just what happens to the rain drop, but how it happens. Directions require both a labeled drawing and text.

In the third draft, the directions still include both drawing and writing, but suggest that the drawing should have a number of elements: "Label each stage and as many parts of your drawing as you can."

Both drafts 2 and 3 provide students with knowledge that they are describing a cyclical process. As a result, the concept that evaporation and condensation occur cyclically is not being assessed. If the task were being designed to assess that understanding, the directions should be revised again to eliminate references to "cycle" or "when you become a raindrop again."

Sometimes the language used in an assessment task, although familiar to its adult developers, can result in performances which may obscure what students really know. An example comes again from Redwood City School District where a prototype task asked

A Work in Progress

Figure 7
Portfolio Contents
O'Farrell Community School
San Diego City Schools

1990-91 Portfolios

- pieces from each subject area (math, science, social studies, language arts, computers, and P.E.) which demonstrate growth
- best work, to date
- a group effort
- something creative
- something the student would like to work on again

Revised 1991-92

- items that show social/ emotional and academic growth
- a multi-media project
- items that show meaningful work in all content areas
- wild card items selected by student
- evidence that essential questions have been answered
- thematic unit project evaluation
- community service documentation
- student / peer / parent portfolio reflection

A WORK IN PROGRESS

Figure 8
Progressive Revisions
Open-Ended Science Questions
5th Grade Science
Chula Vista Elementary School District

Version 1:
Topic: Rain
Imagine you are a falling raindrop. Describe how you became a raindrop and what will happen to you now. Be sure to include pictures with labels to show your adventures.

Version 2:
Imagine that you are a falling raindrop:

1. Tell how you became a raindrop and how, after you fall, you will become a raindrop again.
2. Draw pictures to show the stages you go through.
3. Label your pictures.

Version 3:
Imagine that you are a falling raindrop. Think about how you became a raindrop and how, after you fall you will become a raindrop again. Then describe this in writing on page 2. On page 3 make drawings to show each stage in your cycle. Label each stage and as many parts of your drawings as you can.

REMEMBER:

- Think before you begin to write or draw. You may want to list your changes first on the back of the booklet. (page 4)
- Check to see that your drawings match your writing.
- Be sure that your drawings are well labeled.
- Do your best work.

students "What conclusions can you draw from the story?" Taking the directions literally, one student produced a picture.

Determining Strategies for Administration of Assessments

The potential of a performance assessment to influence teaching and learning can be greatly affected by how teachers manage or administer the assessment. Directions to teachers and students must be clear so that all students are given a full opportunity to perform well. The tasks themselves must be structured so that teachers can integrate them into their instructional program in ways that enhance, not inhibit, learning.

Instructionally sound assessments blur the lines between assessment and instruction, but in doing so, raise numerous questions about administration and management. These generally include, but are not necessarily limited to

- How much help can the teacher give as students do the assessment?
- Should tasks be scaffolded to provide some common preparation/prior knowledge for all students?
- How should teachers balance the time for managing assessment and providing instruction?
- What resources, if any, should students have accessible when they do an assessment?
- What time limits, if any, should be imposed on students to finish an assessment?

The ways in which CAC pilot projects answer such questions are unique to the purpose and the technology of the assessment they have developed, but some strategies found useful in several projects include building common preparation activities into the task, basing decisions about resources on desired outcomes, and making records of differentiated assistance.

Building common preparation activities into the task - The San Lorenzo Valley, Chula Vista, and Redwood City district projects use on-demand, language-rich assessment tasks. In order to assure some commonality of prior knowledge and resources, these projects have built unscored pre-assessment activities in some or all of their tasks. One familiar pre-assessment activity is to have the class brainstorm a list of ideas or vocabulary related to the task the students will be asked to perform. The resulting word-bank is then made available to students as they perform the related task. Another strategy is to structure some form of small group work as a pre-assessment activity. The ACE K-2 Project includes in each performance task an "instructional set": prompts designed to explain what should be expected in a unit of preparatory instruction.

Basing decisions about resources on desired outcomes - Decisions about which resources to make available to students should be based on a clear understanding of what the task is meant to assess and a sensitivity to the fact that all students must have equal opportunities to demonstrate their knowledge and abilities. For example, the question of whether or not to allow the use of calculators on a math assessment should be based on whether the assessment is designed to measure the ability to accurately carry out an algorithm or the ability to strategize solutions to complex multi-step problems in which accurate calculations are only one factor in success. Similarly, a decision to allow calculators implies a responsibility to assure that all students have access to them.

When the leadership of the Mt. Diablo math assessment recognized that students received inequitable opportunities to perform due to differences in the amount of assistance given, the district developed assessment administration workshops for all teachers.

Making records of differentiated assistance - Assessment strategies which rely heavily on observations and anecdotal records provide a unique opportunity to provide assistance and instruction while monitoring an assessment activity. Teachers using the CLR and participants in the ACE K-2 Project are encouraged to provide whatever resources or assistance a student requests, but to record differentiated assistance so it can be considered when performances are interpreted or scored. Teachers practice recording not only what the student does, but also the context which provides clues to her ability to perform independently. Observational records for two students who turned in similarly well-organized essays might differ as follows:

> Student A — *"after I worked with J. to extract main ideas and evidence, she was able to redraft her essay"*
>
> Student B — *"S. first draft was organized by main idea and supporting evidence"*

Developing an Interpretation/Scoring Framework

Decisions about the formats, scales, and dimensionality of the interpretation or scoring of an assessment and about the levels of performance expected can greatly affect its impact on instruction. Within the CAC pilot projects which have considered how to interpret or score student performances, we see the following strategies for interpretation or scoring:

Descriptive Interpretation - In this strategy the assessor describes the student's performance in comparison to a stated standard using narratives or anecdotal records but no single score. Descriptive interpretations lend themselves well to assessments designed for the purpose of describing what students know and can do but are difficult to use to compare a given student's performance to others.

In the Oceanside GATE portfolio project, teachers label student work samples with descriptions of the ways in which that work demonstrates exceptionality. Content standards for exceptionality are included as part of the portfolio format. The teacher presents the portfolio orally to the GATE identification team which recommends for or against nomination to the GATE program.

In the ACE K-2 Project, teachers record their interpretations of a student's performance in terms of how it matches or differs from the stated narrative performance standard. The interpretation guide (Figure 9) is used as a record of the teacher's analysis and plan for future instruction.

Uni-dimensional Rubric - A rubric describes the characteristics of a given performance at several levels of accomplishment. Uni-dimensional rubrics are used in holistic scoring to assign one overall score to student work. Assessors consider the response as a whole, not its component parts, to arrive at a rating. To assign a rating to student work, scorers are guided by both narrative descriptions of the characteristics of each score and anchor papers or exemplars. Uni-dimensional scoring provides a means of assuring comparability of scores across different tasks.

The Mt. Diablo Curriculum-Based Assessment in Mathematics uses a single uni-dimensional rubric developed for each task. Progressive drafts of one such rubric are illustrated in Figure 10.

Multi-dimensional Rubric - This framework is similar to the uni-dimensional rubric, except instead of a single score for a content area, the content area is

A Work in Progrress

Figure 9
Excerpt from Interpretation Guide
ACE K-2 Alternative Assessment Project

Student Name: ______________________

Does the child use strategies to make herself understood and to derive meaning from what others have to say?

IN A STRONG PERFORMANCE ON TRY 3: There is evidence that the child reads for meaning and can call upon a large repertoire of strategies to make meaning from text she does not understand. The child can call upon prior knowledge, use all three cuing systems (graphophonic, semantic and syntactic), paraphrase, use picture clues, vary reading speed, and use other strategies to make meaning from text she reads, or has read to her, which she does not understand. She sees or hears patterns in text or speech and can use the patterns to predict and make the language meaningful. When faced with unfamiliar text, the child can call upon a variety of appropriate resources for assistance. There is evidence that the child contributes to and makes use of pre-writing activities conducted in class. She may begin to initiate pre-writing activities to generate ideas before writing or overcome a writer's block. When prompted to do so, the child will revise the first draft of her writing in order to improve it in some way. With assistance from the teacher, the child edits her writing accurately. The child may use writing to explore or organize ideas even when she is not required to do so. The child uses strategies to assure that she is understood and that she understands what she hears. There is increasing evidence that the child thinks about or plans what she has to say. Without prompting she will seek clarification or consult resources to make sure that she understands what has been said or read to her.	Look for more evidence of: Teaching and learning goal:

broken down into several distinct dimensions, each of which is rated separately. This type of scoring is typically used when finer distinctions are desired and assessors believe that students are likely to be at different levels of achievement on different dimensions.

The writing portfolio project at Stevens Creek Elementary School (Cupertino Union School District) provides an example of a multi-dimensional rubric through which students receive feedback about the aspects of narrative in which their work shows strength and those in which it could be improved. The rubric includes separate descriptors of performance related to the development of character, setting, plot, and theme as well as the student's use of the writing process, attention to conventions, and ability to communicate with the reader. An excerpt from this rubric is shown in Figure 11.

Descriptive Interpretation/Several Uni-dimensional Scales - These assessments combine descriptive interpretive notes with a summary rating according to a single scale for each area rated (e.g. reading, writing). For instance, in addition to their descriptive commentary, secondary teachers using the California Learning Record (CLR) summarize their descriptions of student reading achievement on a five point scale

Figure 10 Progressive Revisions Math Scoring Rubric Mt. Diablo Unified School District

1990 Competency Levels	Rating/ Score	1992 Competency Levels
	6	**Exemplary Response** - complete response with a clear, coherent, unambiguous, and elegant explanation. - includes a clear and simplified diagram when appropriate - communicates effectively to the reader - shows understanding of the open-ended problem's mathematical ideas and processes - identifies all the important elements of the problem
	5	**Competent Response** - fairly complete response with reasonably clear explanation - may include appropriate diagram - communicates effectively to the reader - shows understanding of the open-ended problem's mathematical ideas and processes - identifies the most important element of the problem
Exemplary Response -systematic or elegant approach - organized recording system - complete and accurate response - clear and thorough explanation	4	**Minor Flaws But Satisfactory** - completes problem satisfactorily but the explanation may be muddled. - argumentation may be incomplete - diagram may be appropriate or unclear - understands underlying mathematical ideas - uses mathematical ideas effectively
Satisfactory Response - systematic approach - organized recording system - reasonably accurate response - clear explanation	3	**Serious Flaws But Nearly Satisfactory** - begins the problem appropriately but may fail to complete or may omit significant parts of the problem - may fail to show full understanding of mathematical ideas and processes - may make major computational errors - may misuse or fail to use mathematical terms - response may reflect an inappropriate strategy for solving the problem
Inadequate Response - inaccurate mathematical response - explanation unclear	2	**Begins, But Fails to Complete Problem** - explanation is not understandable - diagram may not be clear - shows no understanding of the problem situation - may make major computational errors
Off Track - significant parts of the problem omitted - explanation incomplete or not attempted	1	**Unable to Begin Effectively** - words do not reflect the problem - drawings misrepresent the problem situation - copies parts of the problem but without attempting a solution - fails to indicate which information is appropriate to problem
No Attempt	0	**No Attempt**

BEST COPY AVAILABLE

A WORK IN PROGRESS

Figure 11
Excerpt from Initial Draft of Rubric for Narrative Writing
Stevens Creek School, Cupertino Union School District

Plot:

random ⟷ sequential   static ⟷ conflict

I. One or two events with little or no conflict

II. Beginning sequence of events, but occasional out-of-sync occurrences; events without problem or problem without resolution ("Once there was a cat. The cat liked milk. The cat ate the rat.")

III. Single, linear episode with clear beginning, middle, and end. ("Once upon a time there were two friends named Frog and Road. Once when they were tree climbing, Frog got stuck. Toad helped him down and they lived happily ever after.")

IV. Plot increases in complexity with more than one episode; episodes contain four critical elements of problem, emotional response, action, outcome; beginning relationship between episodes.

V. Stronger relationship between episodes (with the resolution in one leading to a problem in the next); beginning manipulation of the sequence through foreshadowing, and subplots.

VI. Overarching problem and resolution supported by multiple episodes, rich variety of techniques (building suspense, foreshadowing, flashbacks, denouement) to manipulate sequence.

Writing Process:

arbitrary ⟷ planned   one attempt ⟷ revision

I. Writer works "in the moment" (with one attempt of short duration); planning centered on drawing and talk

II. Child may take advantage of teacher-led group planning (by copying) but will not stray far from the forms offered; revision focused on convention rather than communication.

III. Writer makes more consistent use of group planning, and begins to expand on forms offered (adding new words and concepts to those on the board); revision still centers on convention with few changes in content, structure, voice.

IV. Planning expanded to include organizational structures, but these may not be referred to once the writing begins; during revision writer makes some communication changes in the beginning or end of the text, writer begins to make revision at word level to refine and enhance the text.

V. Writer uses prewriting ideas to help guide the story; planning continues to expand to the use of research (drawing, talk, structures, and/or life experiences still valuable resources); revision includes adding or deleting information to the middle of the text to improve the narrative; word level revision continues but connects to overarching choices in metaphor.

VI. Writer's early planning and extensive exploration of narrative possibilities results in higher quality first drafts; revision centers on communication as the writer moves toward meaning (shifting perspectives between writer, reader, creator, and critic); may make major alterations and reorganization as writers come to have higher standards for themselves.

Developed by Wolf / Quellmalz / Gearhart / Whittaker 1992

ranging from "inexperienced reader" to "exceptionally experienced." The scale is used to make a summative judgment of a student's development as a reader, with the developmental labels tied to narrative explanations of progress at each level. (Figure 12) A different uni-dimensional scale is provided for scoring each genre of writing to be included in the CLR Secondary Language Arts Portfolio.

Task-based Analytical Scoring - Each task in the assessment is rated individually and assigned points which indicate the number of correct or critical ele-

A Work in Progress

Figure 12
Reading Scale from California Learning Record

EXPERIENCE AS A READER ACROSS THE CURRICULUM: SECONDARY

INEXPERIENCED

Inexperienced Reader 1	With limited experience as a reader, this reader generally chooses to read easy, brief texts. Has difficulty with any unfamiliar material. Needs a great deal of support with the reading demands of the classroom. Rarely chooses to read for pleasure.
Less experienced Reader 2	A reader who is developing fluency as a reader and reading certain kinds of material with confidence. Usually chooses short books with simple narrative shapes. Reading for pleasure often includes comics and special interest magazines. Needs help with the reading demands of the class and especially with reading complex literature and using reference and other informational text.
Moderately Experienced Reader 3	A confident reader who feels at home with books and is developing stamina as a reader. Is able to read for longer periods and cope with more demanding texts, including novels and poetry. Willing to reflect on reading and often uses reading in his/her own learning. Selects books independently and can read juvenile fiction and non-fiction and use information books and materials for straightforward reference purposes, but still needs help with unfamiliar material.
Experienced Reader 4	A self-motivated, confident, and experienced reader who may be pursuing particular interests through reading. Capable of tackling some demanding texts, she or he can cope well with the reading in class. Reads thoughtfully and appreciates shades of meaning. Capable of locating and drawing on a variety of sources in order to research a topic independently.
Exceptionally Experienced 5	An enthusiastic and reflective reader who has strong established tastes in fiction and/or non-fiction. Enjoys pursuing personal reading interests independently. Can handle a wide range and variety of text, including some adult material. Recognizes that different kinds of texts evoke different reading stances. Able to evaluate evidence drawn from a variety of information resources. Is developing a critical awareness as a reader.

EXPERIENCED

Published as a component of Portfolio Assessment and Chapter 1: The California Learning Record, a project funded by the Compensatory Education Office, California Department of Education. For further information, call (619)534-4430.

ments included in the response. Typically the ratings for each task are then aggregated to produce a score for the entire assessment. Figure 13 shows an analytic scoring rubric for a reading task from Redwood City's Performance-Based Assessment Project.

Holistic and multi-dimensional rubrics are the strategies most often employed in assessments designed to emphasize the use of prior knowledge to construct meaning in complex tasks. Analytic scoring frameworks run the risk of suggesting that complex intellectual tasks and processes can be reduced to the identification of a few key elements or the demonstration of isolated skills.

The choice of a scoring/interpretation framework can have considerable impact on the amount and quality of feedback provided to students and teachers. Those with the greatest potential to impact instruction are those which provide the greatest amount of informa-

A WORK IN PROGRESS

Figure 13
Prototype Scoring Guide
Redwood City Performance Assessment

GRADE 5	**TASK 1 "In the Shadow of the Eagle" and "Eagles"**
Activity 4	**Reading Standard Assessed:** Interpret text by applying critical/evaluation skills - understand character development and behavior. (II. B. 5. f)

Exemplary Response	Student may choose any of the three words. Work must be exemplified from text.
Brave:	"I put she was brave because the eagle could have killed her. And also because of the mother of the calf found the calf the mother would attack Anna." "She saved an animal by scaring off a big eagle. She carried the reindeer to her tent."
Caring:	"I chose caring because she wanted to save the poor reindeer calf." She was willing to carry the calf home and care for it." "I think Anna was caring because she took the baby deer back to camp with her family to fix it up and stop the bleeding when it got hurt from the big eagle."
Frightened:	"I chose frightened because the eagle was swooping down on her and scaring her. Also while she was carrying the calf, she kept thinking that the eagle could appear again and hurt them both."
1 Response:	"I chose caring because in parts of the story she was very caring for the calf." (overly general)

SCORING

Points

2=	synthesis of multiple ideas or one idea expanded
1=	one overly general or overly specific idea
0=	other, e.g., "she was nice and brave"

Developed by district development team with assistance from CTB MacMillan/McGraw-Hill

BEST COPY AVAILABLE

tion about what the student did well and what errors or patterns of errors were made. Unfortunately, this level of detailed feedback is cumbersome and more difficult to aggregate for the purpose of making comparisons among students or reporting for the purposes of accountability. In making decisions about an interpretive or scoring framework, assessment developers must consider the purpose and audience for their work and the sources of student achievement data various audiences already have at their disposal.

Determining Performance Standards - How Good is Good Enough?

The choice of an interpretive or scoring framework must be followed by efforts to negotiate and build a consensus around the standards which will be used to judge student performance. When teachers share a common understanding of the standards on which student work is judged, they are better able to give students information which they can use to improve their work.

Most CAC projects which have already developed scoring processes and standards use similar processes for developing rubrics. Typically, participating teachers work together using a large number of student work samples. The samples are sorted into groups of generally "strong," "typical," and "weak" work. The first draft of the rubric is developed by describing the general characteristics of papers in each category. Decisions about the range of the point scale are defined by the gradations of quality the rubric writers wish to describe. Most rubrics use a six-point or four-point scale.

Once begun, the conversation about standards inevitably stumbles over the degree to which the standards should be explicitly stated. Is it enough to say student work should demonstrate "attention to detail"? Or should we describe the type and amount of detail we expect. There is little doubt the latter provides students and teachers better guidance for planning teaching and learning, but capturing and agreeing on the former can be a significant first step.

Another consideration comes even after explicit standards have been spelled out in rubrics. Although the rubric describes the characteristics of student work at various levels of performance, it does not necessarily complete the answer to the question, "How good is good enough?" unless a cut-score is established. A cut-score names the level at which students are doing acceptable work. The revised rubric from Mt. Diablo's math assessment in Figure 5 establishes a "4" (Minor Flaws but Satisfactory) as the cut-score.

At this early stage in their development work, most rubrics designed in CAC projects do not establish a cut-score. This may be because participants have yet to solve the very real problem a cut-score evokes. What opportunities can we provide to students whose work is not yet deemed acceptable? Will students have opportunities to redo their work? To be retaught? To try again? What classroom and school structures will we need to design in order to provide these opportunities?

Developing a Procedure and Training for Reliable Interpretation/Scoring

The reliability of a teacher's interpretation or grading of student work is seldom questioned. All too often, individual teachers determine the criteria for good performance in their own classrooms with little attention to whether students understand, or colleagues share, those standards.

Assessments with the potential to improve teaching and learning require the on-going involvement of teachers in the interpretation and judgment of the quality of student work. If students and school communities are to take judgments made by teachers seriously, then efforts must be made to increase the reliability of the interpretations made, or scores given, by teachers.

The two most critical elements in assuring that any two teachers will interpret or score a performance reliably are 1) ensuring that those teachers have similar understandings of the dimension of the task and the standards for performance, and 2) implementing scoring procedures in which assessors focus on the work in relation to commonly accepted performance standards and are not biased by issues such as the incentive to agree with a colleague or to reward effort.

During their first year of development work most CAC pilot projects addressed issues of reliability by identifying and increasing teacher understanding of the characteristics of a good performance. After preliminary trials in their own classrooms, teachers in the Elementary Language Arts and Science Assessment Project in San Diego City Schools were encouraged to develop and revise their own rubrics and to share them with colleagues. The Math/Science Assessment Project in Chula Vista School District provided teachers with a generic rubric and encouraged individuals to practice scoring their own students' work using it.

The GATE portfolio project in Oceanside School District and the ACE K-2 Project conducted teacher training sessions in which concrete examples of student work were reviewed and discussed. Participating teachers practiced scoring the work and discussed and moderated their findings in the training sessions in an effort to increase the reliability of interpretations of student work when teachers used these assessment techniques in their own classrooms.

In the spring of 1992 both Ben Franklin and Horace Mann Middle Schools in the San Francisco Unified School District, participating in the PACE Portfolio Project, designed audit panels to review student portfolio presentations. The audit panels, which included teachers, district office representatives, administrators, and students, were designed, in part, to begin to assess the reliability of the interpretations of work in student portfolios.

Although not formally confronting the question of reliability through official scoring sessions, these projects are working to prepare teachers to score or interpret student work reliably by working to build a common understanding of high quality work.

Scoring sessions for CLAS on-demand tasks provide teachers opportunities to learn strategies for scoring reliably. In CLAS scoring sessions, participating teachers are oriented to the tasks and then review anchor papers which illuminate various aspects of the previously developed rubric. Scorers are organized into table groups, each led by a table leader. All student work is anonymous. Each work sample is read and scored twice. Second readers are not aware of who did the first reading or what score the work sample was given. Table leaders are responsible for identifying papers with discrepant scores in order to arrange a definitive third reading. Table leaders also monitor scoring and retrain scorers whose scores "drift" or who may give consistently discrepant scores.

The Mt. Diablo Curriculum-Based Assessment in MathematicsProject used scoring procedures similar to those used by CLAS. In order to assure that second readers do not see the scores given by first readers, the first score is written in ink which can only be read under ultraviolet light.

Developing a format for reporting student achievement data

How student achievement data are reported can also influence the potential of the assessment to improve teaching and learning. In developing reporting formats and procedures, consideration must be given to factors such as 1) the purpose of the assessment, 2) the timing of reports, 3) the reporter and the audience for the reports, 4) the amount of contextual information which should be provided along with scores or interpretations of student work, and 5) the form the reporting should take, whether in consultation or conference, in written form, or both. A reporting format which is instructionally sound is one which provides timely reports in a way that can be clearly understood and used by the intended audience.

As relatively early experiments in the development of assessments designed to enhance teaching and learning, the primary audiences for the student achievement data in CAC pilot projects have been those teachers involved in developing or implementing the new assessment, parents, and students. Few projects have yet tackled issues of reporting student achievement data to other teachers or to more public audiences such as a School Site Council or Board of Trustees.

The Mt. Diablo School District project helps illuminate the challenge of providing student achievement data which are valuable to teachers. The district's Curriculum-Based Assessment in Mathematics has increasingly shifted its focus from an assessment which might be used for program or school accountability to one aimed directly at affecting change in the teaching of mathematics in the classroom. As a result of that shift, the format, timing, audience and even the messenger for reporting student achievement data has also changed.

In the project's field test year, student achievement data from several different items were aggregated in school-level reports presented by the district's Director of Evaluation and Research at the start of the next school year. In the project's second year, similar school level reports were provided, but this time they included district-wide data, the teacher's aggregated class scores, and representative anchor papers, and were presented in staff in-service opportunities conducted by both the math curriculum support services and the Department of Evaluation and Research.

In its third year, the project was designed to provide teachers with individual scores for students on individual items. The test was given earlier in the school year so that findings could be used to improve instruction in that same year. The teacher's class scores were made available only to that teacher and the principal, who were encouraged to use them to assist in program planning. A public report of findings on the assessment provides district-level results, an analysis of student performance data, representative anchor papers, and recommendations for the improvement of the math program.

One of the biggest challenges in implementing and sustaining support for alternative assessments will be the challenge of communicating the rich but complex student performance data they provide. Parents and community members are accustomed to the data distilled down as letter grades, percentile ranks, or stanines.

The California Learning Record provides an opportunity to consider strategies for sharing descriptive interpretive data with parents. The CLR document itself is a parent reporting instrument. The teacher's narrative and summative assessment on the reading and writing scales is given to parents along with the work included in the portfolio, preferably at a parent/teacher conference. A copy of the Record and the teacher's anecdotal records become part of the student's school file. Although no formal study of the level of parents' response to the CLR as a reporting instrument is yet available, many teachers in the project report parents' pleasure in receiving such a detailed record of their child's work and progress.

Another challenge in reporting student achievement comes in translating performance assessment data to summative reports such as report cards. Several of the districts in which CAC projects are housed have developed report card formats more amenable to new assessment strategies. The La Mesa/Spring Valley and Poway districts have designed report cards that report student progress in Language Arts using developmental scales for reading, writing, and listening/speaking. Figure 14 displays the Reading Stages used in the grade two progress report in La Mesa-Spring Valley. Teachers in both districts have reported that they use their student portfolio entries as evidence for determining which level to mark on the report cards. They have also used the portfolios to share with parents the relationship between their students' work and the report card scales.

O'Farrell Community School (San Diego City Schools) has developed its own report card to match the school's standards for students. Even though the district has a common report card format, the school has received a waiver to pilot and use a reporting framework more aligned with the school's philosophy, curriculum, and instruction. The report card includes a checklist for personal and social responsibility, space for narrative reports on physical education, music, foreign language, industrial arts, computer, and arts performance, and a rating scale for quarterly thematic unit outcomes in technics (math and science), humanities (language arts and social studies), and class participation. Teachers also rate the quality of the contents of student portfolios from the quarterly thematic unit and include comments about any aspect of student performance.

Challenges, Cautions, Caveats

The Challenges of Developing Quality Tasks in Schools

Teachers at work in schools today generally understand assessment through the models which have been provided to them — standardized tests or classroom assessments which measure the correctness of isolated facts or skills. As teachers develop alternative strategies they must challenge current practice and undo many well-ingrained understandings. Observations across CAC pilot projects suggest that there are many questions to be answered about what makes a meaningful assessment task. Teachers are quick to invent tasks which actively engage students in demonstrating what they know, but are less certain when it comes to the challenge of designing those tasks around what are often referred to as "the big questions" or essential elements of a discipline. It is not at all uncommon to see teacher-developed performance tasks which assess one discrete skill or a set of isolated skills.

The assessment tasks and scoring standards teachers develop mirror teachers' knowledge of the content and their instructional expertise, and can only be as good as that knowledge. Efforts by teachers to make assessment more curriculum-based can often mean that the new assessments mirror weak classroom instructional practice. Site-based assessment development efforts can benefit greatly from collaboration or consultation with curriculum or content area experts.

A Work in Progress

Figure 14
Grade 2 Student Progress Report
La Mesa — Spring Valley School District

Reporting Period
(Indicate current
stage with an X)

Academic Areas: Language Arts

Reading Stages

1	2	3	4	Stage 1: Minimal experience as a reader. Has difficulty in selecting material for use. Consistently needs support with reading.
				Stage 2: Has limited experience as a reader. Chooses to read very easy and familiar material. Has difficulty with unfamiliar material. Needs support with reading. Rarely chooses to read for pleasure.
				Stage 3: Is developing fluency and reads short books and simple narratives with confidence. Enjoys favorite or familiar books. Needs help with reading in the other subject areas, especially using reference and information books.
				Stage 4: Reads fluently and with expression. Selects appropriate materials from other subject areas and begins to use reference books. Summarizes, makes predictions, and begins to draw conclusions.
				Stage 5: Is self-motivated, confident, and explores interests through a variety of reading resources. Capable of reading in the other subject areas. Analyzes materials to draw conclusions and make assumptions.
				Stage 6: Is an enthusiastic and independent reader. Reads a wide range of materials. Is able to evaluate and analyze evidence drawn from a variety of sources.

The Temptation to Take Short Cuts Which Emphasize Product over Process

The development and implementation of assessments with the potential to improve learning is time consuming work. There are many places where school practitioners will be tempted to cut corners. The challenge is to think of "assessments" not as sets of tasks, but as the professional process of analyzing student work for the purpose of guiding future instruction. Thought about in this way, decisions about the essential elements in assessment development and implementation are then guided by the potential effect on instructional practice.

One common temptation is to draft scoring rubrics at the same time tasks are initially written, skipping the refinements made in response to student work samples. Yet our observations in several projects suggest that it is the act of revising the rubric up against actual student work samples which provides the most significant insights into the ways expectations for student performance are best described. Where early drafts of rubrics might suggest a standard for "well-organized" work, revisions after analyzing large numbers of student samples spell out specific criteria for organization such as "states problem and follows with supporting evidence," "has distinct beginning and closure," or "uses headings and sub-headings to label ideas." It is in having arrived at these specific criteria that teachers learn to communicate information to students which can be used to improve their work.

Caution: Creating New Paradigms in Old Structures

Many believe assessment reform is the linchpin to real change in school design and practice. Yet most alternative assessment efforts are developed by those immersed in the constraints of current practice. Are we to develop assessment alternatives which guide the structuring of schools we have yet to see? Or limit

BEST COPY AVAILABLE

ourselves to what we can use in the schools we know today?

As we shall see in Chapter 4, teachers worry about the time involved in assessment development and the time and expertise required for successful implementation in the classroom. Certainly administrators and policy makers are concerned about the costs of new assessment strategies. And we must recognize the public information and education effort which will be required to support and sustain assessment reform.

The CAC's projects suggest that we ought not to limit our vision to what may be currently feasible. Many of the CAC's projects are outgrowths of, or have resulted in, major restructuring efforts. Teachers are rethinking the ways they organize time and instructional activities to free themselves to observe students working in small groups or to confer with individuals. Administrators are rethinking daily schedules to create longer blocks of time that can accommodate interdisciplinary or thematic units of instruction and the investigations or exhibitions designed to assess them. School personnel and members of the community are being used in new ways to help students prepare for and demonstrate their ability to do meaningful work. Only by creating assessments which push on, and strain, the current structures of schools will we begin to see what may be possible.

A "6" Today May Not be a "6" Tomorrow

As described earlier in this section, rubrics should be developed using actual student work samples. To do otherwise is to miss opportunities for teachers to negotiate standards and learn to state them explicitly. But because our students today have so little experience responding to performance tasks, we should remember that the best work samples we see in initial trials may be only the upper mid-range samples after students have had opportunities to use the rubric to better understand how to improve their performance. We must expect to, and make opportunities for, raising our stated expectations for teachers and students.

Just as limitations in school structures should not limit the possibilities for the format, interpretation, and scoring of alternative assessments, students' limited familiarity and success with early trials of performance assessments should not cause us to lower our expectations for how they might perform in the future.

The Need to Monitor Fairness and Equity

No assessment is fair unless students have equitable opportunities to perform well on it. Assuring fairness and equity requires that assessments not only be free of bias, but that all students have access to the time and resources necessary to prepare to perform well.

Standardized, multiple-choice tests shortchange many students. Such tests tap only a narrow range of intellectual and cognitive styles. Test questions and formats have been determined to favor some ethnic, socio-economic, or gender groups over others. Although the very nature of many performance assessments allows more varied opportunities to demonstrate achievement, alternative assessment developers should be vigilant lest their efforts result in the same limitations or bias complaints directed at traditional assessments. Bias reviews conducted by panels of community members, university researchers, and subject matter specialists sensitized to the appearance of bias can help developers avoid designing tasks, formats, or scoring procedures that are biased against children of different cultures, experience levels, genders, or language backgrounds.

But eliminating bias in the construction of the assessment task or scoring procedure cannot assure fairness and equity. Assessment projects must also monitor the opportunities provided to students to prepare for the assessments in which they participate. The concept of determining whether students have had the opportunity to learn material on which they are assessed burst onto the national policy agenda early last year with the release of the final report of the National Council on Education Standards and Testing. That report called for voluntary national standards in school subjects and a related system of assessments. It also recommended the adoption of "school-delivery standards" (now called "opportunity to learn" standards) to assess a school's capacity and performance in enabling students to meet the standards. California has worked with the National Governor's Association to explore what these school delivery standards should be.

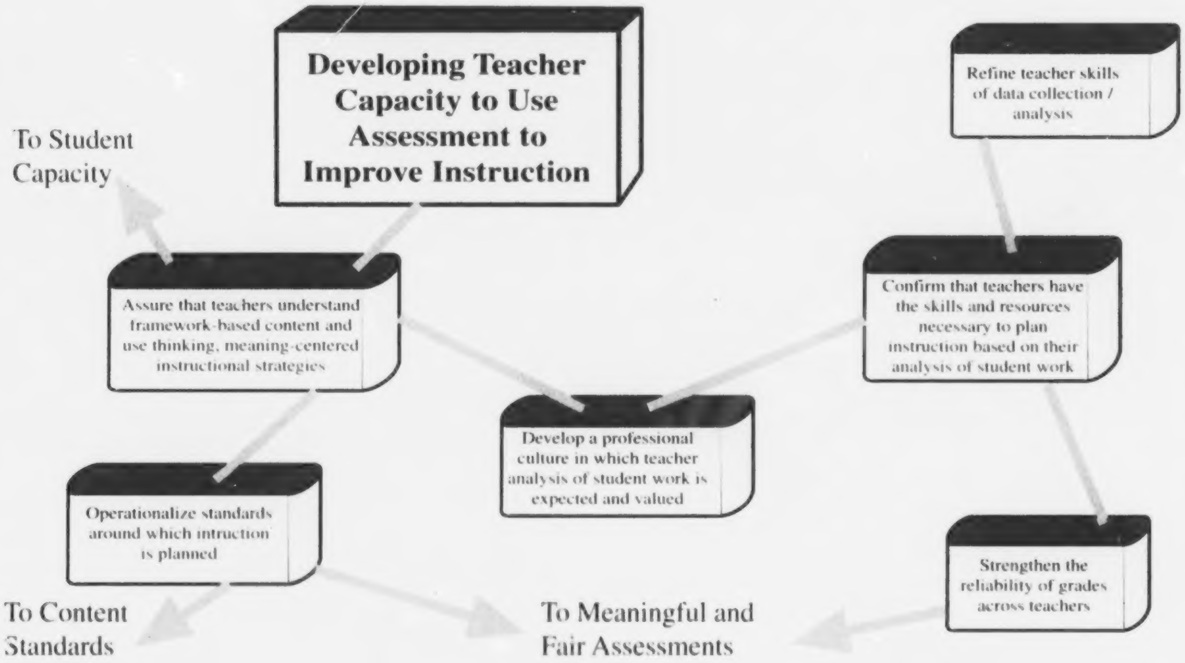
Assessment development efforts embedded in school practice provide a unique chance to consider the opportunities students need, and have had, in order to prepare for assessment tasks. Student difficulty with performance tasks should signal us to look at the opportunities and resources they had to learn and prepare. Published results of the 1991-1992 math assessment in the Mt. Diablo School District make this point most clearly:

> *We hypothesize that the inadequate responses of a large number of students and the apparent frustration of some students may have occurred because students are not accustomed to writing about mathematics. Student performance will likely improve if students are given more opportunities to describe their thought processes orally or in writing while solving problems. Of particular value will be opportunities to solve problems that involve reasoning, recognizing patterns or relationships, analyzing situations, formulating mathematical problems, selecting alternative mathematical strategies, making generalizations, and verifying and interpreting solutions.*
> (Mt. Diablo Unified School District, 1992)

In the next sections we will look at the ways other CAC pilot projects have addressed increased opportunity to learn by building the capacity of teachers and students to use assessment to enhance teaching and learning.

DEVELOPING TEACHER CAPACITY TO USE ASSESSMENT TO IMPROVE INSTRUCTION:

WILL TEACHERS HAVE THE RESOURCES AND SKILLS TO USE ASSESSMENT TO IMPROVE INSTRUCTION?


Developing Teacher Capacity to Use Assessment to Improve Instruction
To Student Capacity
Refine teacher skills of data collection / analysis
Assure that teachers understand framework-based content and use thinking, meaning-centered instructional strategies
Confirm that teachers have the skills and resources necessary to plan instruction based on their analysis of student work
Develop a professional culture in which teacher analysis of student work is expected and valued
Operationalize standards around which intruction is planned
Strengthen the reliability of grades across teachers
To Content Standards
To Meaningful and Fair Assessments

In our enthusiasm for creating performance assessments, we must be vigilant in assuring that those assessments actually contribute to improved performance. If new assessment systems are to have any hope of positively affecting student performance, teachers must have the capacity to use them effectively. Portfolios of student work are little more than static work folders if teachers do not have the skill or time to read and analyze them and to use what they find to plan appropriate instruction. Performance tasks might be more engaging than multiple choice items, but no more valuable in helping students learn unless teachers have the resources necessary to make use of the data available in student work.

Assuring the capacity of teachers to use assessment to improve instruction requires that teachers

- have a common understanding of the standards around which instruction is planned,
- have a solid understanding of both the content they are responsible to teach and effective instructional strategies,
- work in a professional culture in which their analysis of student work is expected and valued, and
- have the skills and resources necessary to plan instruction based on their analysis of student work.

A Work in Progress

Figure 15
Excerpt from GATE Nomination Form
Oceanside Unified School District

Student Profile

Editors Note: Teachers are instructed to check characteristics which the nominee possesses to a high degree, and to provide specific evidence in the GATE Identification Portfolio (of those characteristics).

CHARACTERISTICS: **SPECIFIC EXAMPLES:**

__________ **Potential for extraordinary intellectual development** May be evidenced by: vocabulary beyond age or grade level, advanced skill in written or oral expression, advanced expression of thoughts and opinions, quick and accurate recall of factual information, exceptionally high scores on measure of aptitude, a storehouse of information on a variety of topics, ready recall of visual or auditory information, resourcefulness, knowing where to find answers, curiosity, asking many questions.

__________ **Creative thinking ability** May be evidenced by: generation of many ideas, questions, suggestions, flexibility, readily adapting to new situations, a great deal of imagination, elaborating on and making changes in ideas, concern with improving or adapting objects and systems, unique, clever responses to questions, intense concentration on tasks of interest to her/him, loss of awareness of time when involved in such tasks, appearance of inattentiveness (seems withdrawn, daydreams), impatience with routine tasks, enjoyment of playing with ideas just for fun, enjoyment of simulation, creative problem solving, improvisation, ability to handle a multiplicity of variables, unique, unconventional solutions to artistic, musical, dance, or dramatic problems.

__________ **Leadership ability** May be evidenced by: strong communication skills, ability to get ideas across effectively, direction of group efforts, domination of situations, pleasure in organizing people and structuring activities, respect of peers, enjoyment of other people, high energy, keen insight into others, self-confidence, ability to make decisions based on assessment of problem, ability to carry out plans, ability to gauge appropriateness of timing of decisions, directions, suggestions, patience with others, interpreting for others, ability to inspire others, unusual sensitivity to others.

Reaching Common Understanding of Standards

A standard becomes a useful instructional tool only when teachers know what it means and can identify a variety of behaviors which indicate achievement or progress toward that standard. If teachers do not know what it means for a reader to "challenge a given text," then they are most certainly at a loss for planning learning experiences which will help students learn to do it. If teachers do not recognize the overt behaviors which indicate a student's increasing ability to "challenge text," then they cannot know what that student can do, nor what the student needs to learn. Teachers must have opportunities to make meaning from an abstract standard by putting it in their own words and/ or attaching it to familiar exemplars of student work.

BEST COPY AVAILABLE

The CAC's pilot projects offer some promising strategies for helping teachers make meaning of agreed-upon content and/or performance standards.

Becoming Familiar with Indicators of Performance

One strategy for illuminating the meaning of content standards is to identify a list of performance indicators which suggest progress toward or achievement of a standard. Several pilot projects developed and provided participating teachers with documents which describe specific indicators of student performance. The documents are not used as checklists, but rather as reminders to teachers about the overt behaviors to which they will want to be sensitive.

The Oceanside School District began its Gifted and Talented Education (GATE) portfolio project in hopes of finding a more reliable means of identifying GATE students than offered by standardized test scores. One premise of the project is that giftedness can be demonstrated by more than academic achievement and that portfolios might make it possible for students to demonstrate exceptionality in more than one way. The district GATE portfolio development team provides teachers with instructions for compiling the portfolio and a detailed list of characteristics which may suggest potential for extraordinary intellectual development, creative thinking ability, potential for artistic development, insatiable intellectual interests, abstract reasoning ability, capacity for self direction, leadership ability, and capacity for high academic achievement. Teachers are charged with compiling evidence of some, but not all, of these characteristics in the portfolio. An excerpt from the GATE Program Nomination form which includes these characteristics is shown in Figure 15.

To further teacher understanding of the standards for giftedness, the district invested in a unique partnership. Three school psychologists volunteered to team-teach with six general education teachers for an entire summer school session when the portfolios were first used. Working along with classroom teachers, the psychologists' role evolved into that of instructional and assessment consultants. The partnerships generated conversations about what students' work demonstrated about their learning and how to use the data provided by student work to plan future instruction. Together these professional colleagues came to a more common understanding of the ways exceptionality can be revealed through student work and were able to enrich the portfolio training manual for other teachers by including in it numerous student work samples which serve as exemplars of exceptionality.

Similarly, the ACE K-2 Project includes in teacher binders a list of the observable language behaviors which may serve as indicators of progress toward the content standards (phrased in the form of an essential question) for the assessment. The introduction to this document suggests its appropriate use:

> *Below each question are some indicators — overt language behaviors — which give evidence of the presence of the linguistic knowledge, skill, or attitude in question. ...This list is not comprehensive and not all students will exhibit each indicator, nor is it likely that a given child will be consistent in the use of these behaviors over time. It should not be used as a checklist designed to indicate mastery. Instead, the list is intended as a guide to help evaluators interpret what they see when observing a child and her work. In examples of student work observed or collected over time, use of some of the indicators, or approximations of them, is evidence of growing control and confidence as a language user. The consistent occurrence of these indicators is evidence of a high degree of control and confidence.*

Indicators for one essential question are illustrated in Figure 16.

Sharing Rubrics Up Front

Another strategy for familiarizing teachers with standards is providing them with the rubrics and student work samples in advance of administering assessments. For example, fourth and fifth grade teachers in Chula Vista School District are provided with open-ended science tasks accompanied by a general rubric and a list of task-specific attributes considered essential to an appropriate response. (Figure 17)

Participation in Scoring

The richest opportunity to internalize standards comes about from participation in collegial scoring of student work samples. Often designed for the purpose of assuring reliable scores, scoring sessions prove equally valuable as opportunities to negotiate a teacher's understanding of standards. Descriptions of promising strategies for setting up collegial scoring sessions are described in the previous section.

Building Teacher Expertise in Content and Instruction

In order to provide all students opportunities to achieve high standards, all teachers must have expertise both in their content and in effective instruction. Common sense suggests that the most efficient way to develop good performance tasks is to call on only those teachers with the best understanding of the subject at hand. But as we shall see in Chapter 4, the experiences of teachers in CAC pilot projects suggest that participation in efforts to develop or implement performance assessment strategies may serve as a catalyst for large numbers of teachers to increase their instructional expertise in the content being assessed. Cited below are some examples of the promising ways in which the development of teacher curriculum and instructional expertise has been addressed as an integral part of assessment development.

A WORK IN PROGRESS

Figure 16
Excerpt from Essential Questions and Indicators
ACE K-2 Alternative Assessment Project

Does the child use strategies to make herself understood and to derive meaning from what others have to say?

As a reader:

Does the child use strategies to derive meaning from print?

Uses one or more cuing systems to predict essential vocabulary and/or self correct:
- Student will use semantic cues, filling in what seems to make sense
- Student will use syntactic cues, filling in with a word reflecting internalized grammar or logical word order
- Student will use graphophonic cues, filling in a word based on recognized letters/sounds

- Uses pictures as clues to text she does not understand

Uses punctuation as a cue to make meaning from text

Uses prior knowledge to make predictions and increase understanding

Varies manner and pace of reading based on purposes, intention and audience

Rereads for information when appropriate

Uses a variety of text aides, charts, graphs, dictionaries, pictures, sub-headings and other reference materials to make meaning from text

Can summarize a passage including all major points and uses this as a strategy for making meaning from text

Makes connections (artistic, verbal, written) between text she has read and situations in own life

Any others?

Does the child recognize patterns in literature/text and use them?

Listens attentively when being read to

Orally predicts repetitive language patterns during shared reading

Makes justifiable predictions about storyline during reading

Responds to stories orally, artistically, dramatically or in writing, including some approximation of the storyline

Can identify the elements of a story including:
- the beginning, middle, and end of a story
- the characters, character traits, and interaction
- the setting and plot of a story
- the details of a story

Orally retells story in sequence

Any others?

BEST COPY AVAILABLE

A Focused Curriculum and Professional Development Support System

After the initial orientation of all teachers in the Redwood City School District to the plans and purposes of the Performance-Based Assessment (PBA), teachers were surveyed about their reactions to the items and concerns they had about future implementation of the assessments. Many teachers expressed concerns about the mismatch between the instructional materials and training they had available and the type of tasks students would be asked to perform on the assessment. The district made the decision to direct a large percentage of the resources allotted to PBA to professional development and support. The result is a multidimensional effort which includes

- the development of curriculum directories which include literature units and math lessons directly tied to district standards and model performance assessments. Student copies of the literature used in the unit are included in both English and Spanish, as are audio tapes of the books. The curriculum directories are made available to every teacher at "Author's Events" in which the organization and uses are described, and for checkout through the instructional materials center.

- support for teacher preparation for PBA provided by one or two on-site liaisons, teachers who have been involved in the development and field tests of items and rubrics. Monthly meetings help site-level liaisons keep up on the development of the PBA and give feedback to the district leadership responsible for carrying out the project.

- a bi-monthly newsletter distributed to all teachers which includes announcements of professional development opportunities related to district standards, updates on issues related to PBA, glossaries of PBA vocabulary, results of development activities and field tests, and calls for necessary instructional support materials or assessment task ideas.

- provision of prototype tasks, field test results, and training forms to all teachers as well as opportunities to preview actual tasks on "Viewing Days".

A Work in Progress

Figure 17
Scoring Guide for "Water Cycle" Task 1
Chula Vista Elementary School District

Ideally students' descriptions should match their labeled drawings. However, if either the description or the drawing has considerably more detail, the student should not be penalized...

Students should show their understanding of the stages of the water cycle: ground water, evaporation, condensation (clouds), precipitation (rain, snow, hail). Student's drawing should show all three stages and be labeled.

Students may begin writing at any stage in the cycle, but they should continue it from that point in the sequence.

When scoring, emphasis should be placed on the student's understanding of the process of the water cycle rather than on scientific terminology.

The Redwood City School District's "All Students Will Succeed" Assessment Project demonstrates a unique coordination of the district's professional development, curriculum development, and assessment resources.

Alignment of Professional Development and New Assessment Strategies

The portfolio at Stevens Creek School in Cupertino was designed to evaluate the impact of new forms of technology on student writing for the Apple Classroom of Tomorrow project. Students keep a "working folder," and after conferring with teachers, make decisions about which pieces of their writing to move to a "showcase portfolio." Teachers surveyed after the second year of work with the portfolios recognized that in order to make fruitful use of the portfolio process to improve student writing, teachers had to be more confident of their ability to talk to students about the specific qualities of good writing.

In the third year of the project, teacher development time was not devoted to further discussions of the portfolio template and process. Instead, the staff turned their attentions toward developing a rubric which would give teachers a common vocabulary for talking about narrative writing as well as help them point out characteristics in literature as models for student writing. A segment of the rubric developed through their collaborative work is reproduced in the previous section. On-going work at Stevens Creek involves use of site-level resource teachers who provide classroom support to teachers learning to use the rubric as an integral part of instruction.

Guidelines for Prerequisite Professional Development

When the ACE K-2 Project expanded beyond the initial development team to involve school teams, participating teachers decided to include an introduction to the teacher training binder. The introduction provides a prerequisite list of the curriculum and instructional knowledge teachers should have in order to successfully implement the assessment strategy and help students perform well. The list includes, among others, diagnosing and teaching reading strategies; teaching the conventions of writing and spelling in context; teaching writing as a process; sustaining literature-based reading programs; and integrating reading, writing and listening/speaking instruction.

The teacher training binder suggests that as teachers begin to use the assessment model, they identify through their own work a professional development plan to pursue learning opportunities in areas where their assessment and instruction skills seem weakest.

Building Norms and Expertise which Support Teachers as Researchers

Traditional images of teachers put emphasis on the act of delivering information or planning lessons to assure the acquisition of knowledge. Less emphasis has been placed on the critical role of teacher as the investigator of what students know, what they need to know, and what they are ready to learn. In order that teachers make maximum use of new assessment strategies, they must hone these skills of research and analysis. Several CAC pilot projects focus on these teacher skills:

Kidwatching

Assessments which depend heavily on teacher observations such as the California Learning Record and the ACE K-2 Project use development and support sessions to share strategies for managing observations and for learning to record anecdotal records. Yetta Goodman first referred to these observation and record keeping skills as "kidwatching." (Goodman et al, 1989) Teachers in the CLR pilot projects in the San Diego City School District and in Lakeside Union School District meet during the weekend or on release time sessions to share their experiences using the CLR. The time is used by teachers to interpret the meaning of the content standards and to discuss what specific student behaviors might constitute evidence of achievement of each. Participants report that these discussions sharpen their ability to observe and see growth and progress.

The ACE K-2 Project assessment procedures were revised several times to facilitate observations and the recording of evidence in anecdotal records. Significant training time is devoted to watching videos of student performance and recording and comparing observations with other teachers. Teachers practice recording observations rather than judgments of student behavior, and learn to collect multiple forms of evidence to support later interpretations of student performance. Several schools within the ACE K-2 Project have restructured to permit teachers opportunities to "kidwatch" along with a colleague or to use instructional aides to provide support of assessment activities.

Student Case Study Portfolios

Portfolios at Newcomer High School in the San Francisco Unified School District are used by teachers as a vehicle for studying the scope and content of the curriculum as received by students and the effect of their teaching on students' language acquisition. Findings from Newcomer's portfolios contributed to the annual evaluation of the HERALD Project in which these same teachers participated. The purpose of the HERALD project is to improve the ability of high school teachers to promote higher levels of oral and written communication.

An interdisciplinary group of teachers met regularly to review the work in the portfolios of one or two students, looking in particular to see what the portfolio would reveal about growth in language acquisition and the quality and quantity of writing challenges students were given across courses. The ongoing in-depth examination of the work of a few students was used to plan instruction and curriculum as the year progressed.

Examples of the rich information revealed from close examination of student work over time is provided in the case study of Alma, a ninth grader at Newcomer, who is profiled in the *HERALD Project Evaluation Report 1992* (Greenleaf). Alma's portfolio contains early efforts which are essentially error free, but employ little more than safe, simple sentences. As her writing progresses she begins to take risks, not only in sentence structure, but in the topics and concepts she chooses to write about. Figure 18 provides an example in which she takes on political and economic issues in her Brazilian homeland and argues her position. In this and later work, she begins to create run-on sentences and demonstrate less control of complex ideas. Yet as the evaluation's author recognizes, "(In using writing to express opinions and take positions), Alma is ...creating a function for her writing into which her language will eventually grow."

Careful study of several pieces of Alma's writing reveal that her misuse of vocabulary is often a result of a habit of making literal translations from Portuguese. After interventions from her teacher on this issue, Alma becomes more conscious of her habit; later works show fewer instances of literal translation.

The portfolio collection provides Alma's teacher with a window on her progress as a writer. Her errors are revealed to be "a benchmark of growth rather than as a betrayal of inadequacies." Analysis of Alma's work over time presents many rich "teachable moments" in which Alma's own work reveals what she is ready to learn.

Building Capacity to Plan Instruction based on the Analysis of Student Work

Teachers face enormous pressures to "cover the material" or "finish the book." Under these conditions, instructional plans are developed as a forced march through a scope and sequence of skills or the chapters of a text, with little regard for what students have actually learned, or perhaps failed to learn, along the way. Alternative assessments offer new windows into the minds of our students, new opportunities to examine what students actually understand. However, these new assessment tools can only improve the quality of student performance if teachers have the capacity to use the data they find there to effectively plan what happens next. Teachers must be given opportunities to analyze student performance data and plan instruction which is appropriate to what students most need and are most ready to learn.

A WORK IN PROGRESS

Figure 18
Excerpts from Student Portfolio
Newcomer High School
San Francisco Unified School District

Topic #1 - About Me
2/3/92 Draft #2

My name is Alma. I am from Brazil. I am 16 years old. I came to U.S. with my family last weekend ago. I go to Newcomer High School. I live in San Francisco because San Francisco is big and beautiful. I like to play volleyball, swim, and race horses. I don't like Physics and Chemistry. I have one dog, Pity, but he stayed in Brazil. I have one brother. His name is Jackson. He goes to Newcomer High School also. I take the bus with my brother. Sometimes I fight with my brother. But now we need each other. I don't speak English very well but I speak portuguese well.

Topic #3
2/12/92 Draft #3

The abandoned children of Brazil

Brazil has many abandoned children they don't have house, food, or clothes. Some children live in orphanages but the most live on the street and the culprits for that are the parents and authority. The parents are guilty because they place the sons in the world but don't have money of to support them. The authorities are guilty because don't make birth control, by then if maked birth control and establish than the families than don't have money don't be possible the have sons. Because occur and than the families have less money and are the ones that have more children. These children don't have school specific for them. I feel than they will need to of them have people to take care, as they need much love, affection, and help.

Encouraging Instructional Planning tied to Assessment Data

"The Assessment Sampler," Redwood City School District's all-teacher newsletter, is used to share assessment data from early field tests and to suggest ways teachers can "teach to the district standards." In the fall issues data on student strengths and weaknesses as determined in the field tests were published for all teachers. Newsletters also include teacher-written descriptions of instructional strategies used to address areas of weakness:

In December, JB's sixth grade class tried one of the formative prototypes, the "how to" speech. The class was making a gingerbread village, and in cooperative groups they constructed their part of the village and then wrote speeches which would explain the process to others. They reviewed a rubric we made, practiced their speeches, and gave them to other classes. We taped their presentations and then, with rubric in hand, the students scored their own performances. The next step will come this spring when the students will use the same rubric, but do an individual speech. With such a good model, we feel the students will be ready to perform to their very best ability.
(Redwood City School District, March 1992)

The Benefits of Planning Backwards

Teachers in the ACE K-2 Project include in their training and development session an exercise in instructional planning referred to as "Teaching While Thinking About the End." Before starting a unit of instruction in preparation for a given performance task, teachers collaboratively map the information and concepts students must know to do the performance well. Units of instruction are based on these maps.

As part of the teacher training for the model, teachers review these maps after collaborating on interpretations of several exemplary performances. Discussions among participants center on the question: "What do these performances collectively tell us about what students may need to be taught in preparation for this performance?" Instructional maps used in teaching future units are revised accordingly.

Challenges, Cautions, Caveats

Building teacher capacity to use assessment to improve instruction involves a serious commitment on the part of all components of the school community. From school leaders must come policies which support the professional role of teachers as judges of what students know and need to learn; equally, time must be provided in the teacher's day to gather and analyze assessment data and organize instruction congruent with the next steps it suggests. From teachers must come the commitment to recapture the professional responsibility to assess accurately and fairly and to make that assessment the basis of instructional decision making.

The principles guiding the development of the new California Learning Assessment System remind us of the importance of building the capacity of teachers to carry out authentic performance assessment at the classroom level. "For assessment to realize its full potential to improve instruction, teachers need to be at the center of the assessment process...Their role as the professional 'keepers of the standards' must be recognized or it will never be properly realized." (CLAS Policy Committee, 1993).

But there are still many hurdles to be cleared.

The Challenge of Developing Content and Instructional Expertise

With the development of the state curriculum frameworks, teachers have been asked to make complex changes in their instruction. The frameworks redefine what is important to understand about content areas and make explicit suggestions for major changes in teaching techniques as well. For elementary teachers, these changes are occurring in six different subject areas. Unfortunately, research (EEPA, 1990) on the early implementation of the Mathematics Framework suggests that many teachers are not well prepared to shift from a teacher-centered to a more student-centered curriculum.

That work concludes that if teachers were treated like the active learners the frameworks wish students to be, then they would be provided with time and assistance to explore the new areas of content knowledge demanded in the Frameworks, to engage in professional conversations exploring assumptions about how children learn in a subject area, and to draw implications for how that subject might best be taught. They would have opportunities to experiment with and evaluate the instructional practices and curricular sequencing called for in the Frameworks. However, these opportunities are extremely rare, and dwindling resources for schools make it unlikely that they will become more available in the near future.

Although as we will see in Chapter 4, teachers involved in assessment development work see it as a vehicle for improving their understanding of content and instructional expertise, sponsors of these efforts must weigh the potentially detrimental impact of implementing assessments which do not reflect solid content knowledge or instructional expertise.

The Challenge of Interpreting Performance for Instructional Improvement

Even well-designed performance assessments are complicated to interpret. Portfolios, performance tasks, and anecdotal records gather a great deal of information about individual students over time. Their interpretation requires sophisticated knowledge of how skills are likely to be revealed in student discourse and products as well as the ability to focus data collection on the most informative evidence and to efficiently recognize its significance.

The student outcome portion of the 1991 HERALD project evaluation study suggests some obstacles to using portfolios for instructional improvement. As part of that evaluation, students and their teachers were interviewed about the student's portfolio to determine program effects. The evaluator found that teachers were less aware than students of the growth the portfolios revealed. In speaking of student learning, teachers generally did not refer to the portfolios for specific examples. Teachers seemed more prepared to speak of student performance on individual assignments than of student growth across an extended time period. The evaluator attributed this lack of use of the portfolios as a whole both to the lack of teacher time to study the portfolios of all their students and to possible gaps in teacher knowledge of the development of student learning over time, concluding, "Looking for student growth in collections of student work and in what students do and say requires the knowledge of what to look for, the diligence to find it, and the patience to wait for it to appear."

The writer goes on to recommend the use of case studies, the in-depth study of a few students' work over time, as a vehicle for teaching teachers to interpret student work. Other studies report that when teachers examine a few students in depth and discuss them with their colleagues, they develop both sensitivity and interpretive skills that improve their ability to look more closely at other students as well (Snyder, 1992)

The conclusions of the HERALD Project evaluation study might well be heeded by others responsible for implementing performance assessments. A portfolio has little potential to improve student performance if the teacher does not have the time or the skill to analyze the student work in it.

Will Teachers Accept Responsibility for Assessment?

The work of the CAC pilot projects illuminates the full range of teacher attitudes toward assessment. Some eagerly await the challenge and the opportunity to play a greater role in assessment; others see alternative methods as an added burden in an already overloaded array of responsibilities. Even in pilot projects initiated by teachers who wanted tools which would help make a better connection between instruction and assessment, some found the notion of using assessment results to evaluate and plan instruction alien to their work.

What will be the profession's official response to the teachers who react to introduction of alternative strategies with "I don't have time to observe my students! I teach," or to teachers' association efforts to seek compensation for collegial scoring sessions?

The Challenge of Restructuring to Build Capacity

Most of the strategies for building teacher capacity to use assessment to improve instruction cited above result from an initiative to use teacher time or resources in new ways. In these relatively small experiments, this reallocation of time and resources amounts to little more than tinkering. But these experiments foreshadow trends in larger alternative assessment development efforts. The new California Learning Assessment System makes use of teachers employed during short periods of the summer to score on-demand assessment tasks and envisions increasing involvement of teachers in scoring student work at the local level as part of the statewide system.

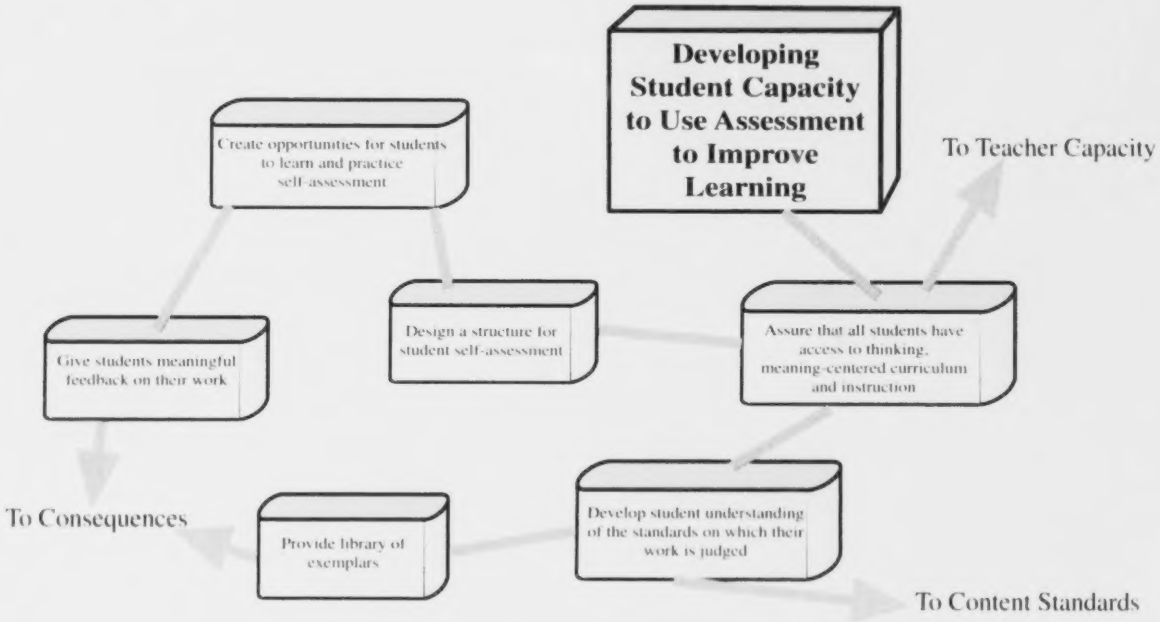
The New Standards Project is training "senior leaders" in participating districts who will train others in assessment development and implementation. Senior leaders will be responsible for training "school leaders" who will coordinate the administration and scoring of exams along with the implementation and analyses of portfolios at the individual school level. These educators will conduct scoring workshops in their schools and help their colleagues interpret the results and their implications for curriculum and school reform. (O'Neil, February1993)

In both these state and national examples we see recognition that if assessment is to reach its full potential to improve learning for all students, then teacher time and responsibilities, and familiar roles and relationships within the school must be restructured. Similar restructuring must be considered if local implementation efforts are to reach their instructional potential.

Unless significant thought and resources are allocated to building the capacity of teachers to facilitate the development and implementation of meaningful curricula and instruction, and to use assessment data effectively, schools will have a difficult time meeting the program delivery standards which will guarantee the opportunity for all students to build their capacity to reach high achievement standards.

DEVELOPING STUDENT CAPACITY TO USE ASSESSMENT TO IMPROVE LEARNING:

WILL STUDENTS HAVE THE RESOURCES AND SKILLS TO USE ASSESSMENT TO IMPROVE LEARNING?


Developing Student Capacity to Use Assessment to Improve Learning
To Teacher Capacity
Create opportunities for students to learn and practice self-assessment
Design a structure for student self-assessment
Assure that all students have access to thinking, meaning-centered curriculum and instruction
Give students meaningful feedback on their work
To Consequences
Provide library of exemplars
Develop student understanding of the standards on which their work is judged
To Content Standards

Meaningful assessment data in the hands of a teacher with the capacity to use it well increases the likelihood of improved instruction. But only when students also have access to meaningful assessment data and the ability to use it will they have the capacity to improve their performance.

School programs designed to assure that all students can prepare for and reach high standards must attend to both the learning resources available to students and to the quality of the opportunities students have to use assessment data to improve their learning. The New Standards Project's (NSP) social compact states:

> *Since, for students, much will depend on meeting high standards of performance, we pledge to uphold our part of a new social compact. Specifically, we pledge to do everything in our power to ensure all students a fair shot at reaching the new performance standards, and to prevent students' performance on the new assessments from being used as the basis for awarding a diploma or any other form of credential unless all students in the jurisdiction awarding the credential have had an opportunity to prepare themselves well. This means that they will be taught a curriculum that will prepare them for the assessments, their teachers will have the preparation to enable them to teach it well, and there will be an equitable distribution of the resources the students and their teachers need to succeed.*
> (New Standards Project)

In *Standards of Practice for Learner-Centered Schools,* Linda Darling Hammond warns that high achievement standards will not lead to better schools unless "delivery standards" or "standards for practice" receive equal attention. Developers of the California Learning Assessment System are experimenting with ways to assess "opportunities to learn" as a part of the new system.

All students must have access to a meaningful curriculum, and opportunities within their instructional program to receive meaningful feedback on their work, to self-evaluate, and to use all the data at their disposal to take responsibility for their own learning.

In the last section, we saw how participation in assessment development efforts provided a catalyst for teachers in several schools and districts to improve their own knowledge of content and instruction. In this section we will look at the strategies CAC pilot projects have used to

- assess the degree of access students have to a meaningful instructional program,
- communicate standards and structure opportunities for their self-assessment, and
- create cultures which nurture self-reflection and personal responsibility.

Assessing Access As Well As Achievement

Several of the CAC's pilot projects began struggling with the issue of "delivery standards" even before the term entered the debate at the national and state levels. These projects saw alternative assessment strategies not only as an opportunity to provide student achievement data, but as a vehicle for looking at the kinds and quality of learning experiences available to students.

Long before the teachers at Benjamin Franklin Middle School in San Francisco Unified began to develop strategies for scoring student performances, student portfolios were used as a window on the "opportunities to learn" provided to students throughout the school. Staff meetings were organized in which teams of teachers looked at a large number of portfolios, not with the intention of scoring them, but in order to pursue answers to questions such as:

> *Are students at our school given enough opportunity to write in disciplines other than English?*
>
> *For what purposes and in what genres do students do most of their writing?*
>
> *Are students challenged to complete complex, long term tasks and projects?*
>
> *What is the quality of the reflection and self-evaluation students do?*

Armed with preliminary answers to these questions, the staff revised the instructional program and restructured resources to enhance the access of all students to the core curriculum.

A Work in Progress

Figure 19
Performance Prompt — Growing Up
ACE K-2 Alternative Assessment Project

We have spent a long time reading, talking and thinking about how people change and stay the same as they grow older. Now you can start your own project to demonstrate what you've learned. You have already done a project which tells about how *you* have changed and will change. This time you will be asked to think not just about yourself, but to report on the life of someone else. You might choose to share the life story of a familiar character, or report on a real person -- someone you know, or someone famous. Use your project to show the many ways this person changes or stays the same as they grow older. Think about their looks, their ideas, their abilities and interests and be sure to use lots of those ideas in your project.

You can do any kind of project which you think will help you share what you have learned, but you must include some original writing that is an example of your best work. When your project is finished, your teacher will ask you to describe it to her or to share it with your friends in class.

Remember that a good project should show that:

1. you used books, people and/or other resources for new information and ideas.
2. you thought carefully about the ideas you got from books, people and other resources, and put them together to make your own original thoughts and ideas.
3. you know and have learned a lot about growing up.
4. you planned something that is interesting and easily understood by others who will see, read or listen to it.

BEST COPY AVAILABLE

The LEP Portfolio project in Poway Unified School District, begun as an effort to develop an instrument to measure student progress in second language acquisition, evolved in its first year into a vehicle for reviewing the kind and quality of instructional opportunities provided to English learners.

Poway's LEP Portfolio development team began their work, not by focusing on a portfolio template, but by sharing instructional strategies which they considered to create valuable opportunities for students to reveal language growth. The sharing, evaluation, and cross-use of these instructional strategies provided teachers opportunities to identify gaps in the instructional program provided to English learners in their own classrooms, and forced reconsideration of the LEP identification and classification procedures throughout the district.

Although much work remains before these projects provide comparable data on student achievement, they already contribute significantly as vehicles for examining the instructional opportunities and practices which support student growth. In Chapter 4 we will see how these internal assessments point the way to beneficial changes in instructional practice and the school program.

Providing Structures And Opportunities For Self-Assessment

Students are the passive recipients of most traditional assessment data. Because traditional assessments are usually administered at the close of an instructional period or unit, both the student's and teacher's attentions have moved on by the time results are available. Even timely results of multiple-choice assessments seldom provide data which can be fruitfully used to improve performance of meaningful work.

Alternative assessment strategies have the potential to involve students in the evaluation of their own work, by teaching them to understand what their work reveals about their achievement and to identify what to work on to improve their performance. The choices made in designing and implementing alternative strategies will have a great deal of impact on whether the assessment can be used in these ways.

Communicating Standards to Students

Students can only learn to assess their own work when the standards for that work are clearly communicated to them. Without clarity on standards, students are likely to think of achievement as a game of chance or luck, resigning themselves to feeling they have little control over the quality of their work.

Many CAC pilot projects have built strategies into their assessments for making standards clear to students. One such strategy involves writing the criteria for success on a given task into the prompt which describes the task to students. Prompts from the ACE K-2 Project and school theme tasks from O'Farrell Community School (San Diego City Schools) provide examples of the ways in which standards can be shared within a task prompt. (Figures 19 and 20)

La Mesa-Spring Valley School District's writing portfolios make students aware of the characteristics of the various writing genres by outlining them on the outside of the portfolio file provided to each student. The outlines suggest a structure or frame for each type of writing, so that students can refer to it as they draft their own pieces. The file also provides suggestions and symbols to guide the students through the editing process. Placed on the front of the portfolio folder,

A WORK IN PROGRESS

Figure 20
Project Criteria
Quarter 1 Project: Presenting Myself
O'Farrell Community School
San Diego City Schools

A. Criteria for Presenting Myself Project

1. Demonstrates knowledge of own physical make-up and dimensions (i.e. contents of body, nutrition, elements and compounds, blood components).
2. Demonstrates knowledge of cultural contributions and geographic origins (i.e. family tree, culture of neighborhood/ family, customs, personal time line).
3. Demonstrates ability to describe qualities of self (i.e., personality, interests, strengths, future foals, learning style, personal journal).
4. Demonstrates pride in work through:
    - a. neatness
    - b. organization of materials (resources), presentation
    - c. completeness
    - d. attention to detail
    - e. timelines in accordance with due dates
    - f. creativity

A Work in Progress

Figure 21
Writing Folder Guidelines
La Mesa — Spring Valley School District

Writing Process

The writing process includes
- Prewriting
- First Draft
- Feedback
- Second Draft
- Final Copy

Editing Checklist

1. Did I indent each paragraph
2. Did I begin each sentence with a capital letter
3. Did I capitalize proper nouns (person, place, thing)?
4. Did I use the correct mark at the end of each sentence?
5. Did I spell each word correctly?

Story

- I. Describe the setting.
- II. Develop the character.
- III. Describe the conflict (problem) and give details.
- IV. Focus on a climax (turning point).
- V. Reach a conclusion (resolution to the problem).

Report of Information

- I. Introduction
  - A. State your topic.
  - B. Narrow topic to specific.
- II. Body
  - A. Support specific areas with details and evidence.
  - B. Use own knowledge and research.
  - C. Express your own involvement with the subject.
- III. Conclusion
  - A. Make a summary statement.
  - B. Leave your reader informed.

Autobiographical Incident

- I. Introduction
  - A. What is the experience you remember?
  - B. When did it happen?
  - C. Who else was involved?
  - D. Where did it happen?
  - E. Why do you remember it?
- II. Experience in detail
- III. Conclusion
  - A. What did you learn from the experience?
  - B. How did you feel about it then?
  - C. How do you feel about it now?

Firsthand Biography Sketch

- I. Introduction
  - A. Who is your subject?
  - B. Reason for choosing subject?
- II. Characterize Subject
  - A. Use incidents, descriptions, and dialogue.
  - B. Reveal relationship to you.
- III. Conclusion
  - A. Why is subject important to you?
  - B. Have summary statement.

these content standards are constantly accessible to students as they write. (Figure 21)

CAC projects provide fewer examples of strategies for helping students learn about the *performance* standards on which their work will be judged or for helping students to understand the gradations in quality of work which separate stronger and weaker performances. This is undoubtedly due to the fact that few projects have yet arrived at performance standards or made them available to teachers. Within several projects, the CAC observed teachers experimenting with strategies which may prove useful in communicating performance standards to students when they are available. These strategies include: sharing rubrics with students at the time assessment prompts or tasks are given, providing a variety of anchor papers which demonstrate performance at various levels, and building informal libraries of exemplars of work which students can analyze in preparation for their own performances.

Providing a Structure for Self-Assessment

The skill of self-assessment is enhanced by good feedback and practice. Assessment strategies which provide a structure for reflection communicate the importance of this skill, as well as provide multiple opportunities to practice it.

One of the most common structures for self-assessment in CAC pilot projects is the use of a reflection sheet attached to pieces selected for inclusion in a portfolio. Students are required to answer a set of questions for each piece of work they include and to attach the reflections sheet to the work in the portfolio. The list of statements and questions in Figure 22 was compiled from the reflection sheets of portfolio projects at Stevens Creek School in Cupertino, Benjamin Franklin Middle School in San Francisco, and O'Farrell Community School and Bell Junior High School in San Diego City Schools. The list was then categorized to show the range of reflections students are asked to do and the ways these reflections can be used to guide improvement of future work.

Informal review of student work across pilot projects shows, not surprisingly, that the quality of the self-assessment done in response to these partial statements and questions varies in proportion to the degree of understanding students have about the standards for the work they have produced. Students who have been given little help in understanding specific criteria for good work often respond with unsubstantiated opinions or relatively unhelpful analyses, such as "It was good," or "I liked it." As students become more informed about what distinguishes good work, they begin to cite specific examples to describe qualities in their work and to make specific plans for improvement.

In the Coordinated Science Portfolio Project at Monte Vista High School in the Grossmont High School District, student self-assessment is structured in two ways. Students are provided task-specific rubrics for assignments within the portfolio in advance of doing the work, and after compiling the portfolio at the end of the reporting period are required to grade their own portfolios before turning them in to the teacher for grading. Students are also provided with a rubric which defines gradations for meeting the portfolio's content standards. (Figure 23) Informal analysis of the reliability between teacher and student scores suggests that building required self-assessment into the portfolio process helps students to more accurately assess their own work and progress.

Providing meaningful feedback to students

The ability to self-evaluate is enhanced by meaningful and timely feedback from others who understand the standards for the work, or who can provide unique personal reactions to it. Several CAC pilot projects have built opportunities for students to receive feedback from third-party observers and parents, as well as from teachers, into their assessment strategies.

Much of the feedback students receive from teachers is oral and remains undocumented. Yet there is little doubt its usefulness in improving student work depends on how well the teacher understands and communicates the standards for good performance. Recognizing this, the teachers in the writing portfolio project at Stevens Creek School in Cupertino have dedicated many of their professional development resources to classroom observation and coaching for teachers. The coaching is aimed at helping teachers use the narrative writing rubric to give students the specific feedback which will help improve their writing.

Students at O'Farrell Community School in San Diego solicit feedback from at least one other person, a student, parent, or other adult, for each entry in their portfolio. That person is asked to fill out a reflection piece which provides the student information on the thoughts his/her work provokes, or what it reveals about her learning. (Figure 24) In presentations where each student summarizes the evidence in her portfolio, students might be asked what the third-party reflections taught them about their work.

Parents at Stevens Creek School in Cupertino are involved in commenting on student work in the portfolio. Portfolios are sent home quarterly so that parents can review the working folder and the student's "showcase" selections. Students have previously completed a reflection form for each showcase piece, and parents are asked to complete a similar form that asks which selection was their favorite and why they liked it. Students and parents often disagree about what is the child's best work, and parent comments reveal that they often gain insight into their child's abilities or personality by viewing the portfolio contents with the child. As we will see in Section 4 this dialogue increases the ability of parents to support the instructional program at home.

Secondary teachers from the San Diego City Schools who are experimenting with use of the California Learning Record engage students in completing much

Figure 22
Questions Compiled from Student Reflection Forms

The following were designed to help students reflect on the purpose of the assignment and the context in which it was given and completed:

One thing I want to tell you about this assignment is...
This selection is most like/least like my work in this category because...
When you read this piece, was it something you had read before, or something that was new to you?
Did you read it alone, with a partner, in a small group or out loud as a class?
Was it assigned reading, or did you choose to read it?

These cause the student to reflect on personal reactions to the work, providing insights into the students' level of engagement and motivation toward this and similar assignments:

When I look at this assignment, I like.... I would improve....
Describe what you did and how you felt during the performance (musical production)
What do you think about the content of the writing (do you like what you said)?
What I liked most/like least about this particular topic/activity was...
What is your overall impression of this piece of writing?

These assess how, or if, the student draws on prior knowledge to complete the assignment:

How does this selection fit into your previous reading (was it harder, easier, different, did it teach you something new, remind you of something you already knew etc.?)
What did you have to know or understand to read this piece well?
What have you done in class this year that helped you in your performance?

These require students to describe the strategies used for carrying out the work

What did you do if you did not understand?
If this was a group activity, what was your particular role in the activity?

These cause students to describe specific learnings gained by doing the assignment:

From this assignment I learned...
How did this kind of reading add to your power as a reader?
How does this piece of writing relate to your overall experience as a writer? Does this piece show off a new skill you have learned, or an improvement you've made in writing? How?

These examine the students' understanding of the standards applied to the performance and the ability of students to assess their own work against those standards:

I have chosen this item in my portfolio because....
I think this paper/project is an example of my best work because...
What do you think about your ability to handle this particular kind of writing (do you like the way you said it)?
If I were to give myself a grade on this assignment it would be a ____ because....
What grade would you give the group for the activity? Why? What grade would you give yourself? Why?

These assess the ability to self-correct and to apply the task to future challenges:

I can apply the information I learned in this assignment to my daily life or other scientific issues by...
If I had the chance to do this activity again, the next time I would...because...
If you were going to do more research/study/experimentation about the (science) in this selection, what would you do?
What experience or support do you need to further develop your reading?
What advice would you give a student doing this performance next year?
What are your goals for next year?
Would you change the way the group worked together? Why or Why not?

Compiled from forms used in schools in Cupertino, San Francisco, and San Diego.

BEST COPY AVAILABLE

of the Record themselves. The student's own observations and reflections are enriched not only by feedback from the teacher but from letters submitted by parents.

Building Learning Cultures

Teaching the skills of self-assessment, though important, will not assure their on-going use. Students must internalize and value the role of self-evaluation in becoming a life-long learner. In order to realize this goal, the schools in which students work should be true learning cultures in which self-assessment is not only taught, but modeled by teachers and other adults in the school community. The school must demonstrate through its activities, symbols, and structures that it values and expects learners to analyze and take responsibility for their own learning.

A good example of some of the ways schools can structure a learning culture can be seen in the systematic efforts to implement portfolio assessment at O'Farrell Community Middle School in San Diego City Schools. Students at O'Farrell are responsible for compiling a portfolio which demonstrates their achievement of the O'Farrell Standard. As illustrated earlier in this section, student self-assessment is structured by providing students the content standards for individual assignments, along with the task description, and requiring students to reflect and seek feedback

A WORK IN PROGRESS

Figure 23
Final Portfolio Check-Off Sheet
Coordinated Science-3rd Semester
Monte Vista High School
Grossmont Union HighSchool District

___ completed cover sheets for each of the following:
___ 2 labs
___ 2 writing activities
___ 1 creative expression
___ 2 gathering information

Cover sheets should be neatly printed, written, or typed and stapled to front of the activity. Activities should be arranged in the order listed above.

Grading will be as follows:

/10 neat

/10 on time

/10 sequenced as listed

/70 Student portfolio includes *all* tasks and shows evidence of:

- growth over time
- creativity
- ability to collaborate in a *positive* manner
- *significant revisions* of work
- *superior understanding* of interdisciplinary themes *and* relevance to Science Technology and Society (STS)

Self-evaluations and coversheets are *thoughtful and documentation is detailed.*

/55 Student portfolio includes *most* tasks and shows evidence of

- growth over time
- creativity
- ability to collaborate in a *purposeful* manner
- *make revisions* of work
- *strong understanding* of interdisciplinary themes *and/or* relevance to Science Technology and Society (STS)

Self-evaluations and coversheets are *thoughtful or documentation is detailed.*

/40 Student portfolio includes *few task types* and shows evidence of

- growth over time
- creativity
- *attempts* to collaborate
- *revisions* of work are minor, insignificant, or poorly documented
- *limited understanding* of interdisciplinary themes *and/or* relevance to Science Technology and Society (STS)

Self-evaluations and coversheets are *vague or poorly documented.*

A Work in Progress

Figure 24
Teacher, Parent/Guardian, Peer Portfolio Evaluation Form
For A Chosen Portfolio Piece
O'Farrell Community School
San Diego City Schools

Student: ____________ Homebase: ____________ Date: ____________

Peer Evaluator:	**Parent/Guardian Evaluator:**	**Teacher Evaluator:**
____________	____________	____________
1. What do you think the directions were for this assignment?	1. What do you think the directions were for this assignment?	1. What do you think the directions were for this assignment?
____________ ____________	____________ ____________	____________ ____________
2. How does this piece show that the student learned?	2. How does this piece show that the student learned?	2. How does this piece show that the student learned?
____________ ____________	____________ ____________	____________ ____________
3. Additional comments:	3. Additional comments:	3. Additional comments:
____________ ____________	____________ ____________	____________ ____________

How would you use the information above to improve this assignment? Use a separate sheet of paper to answer.

from several sources before including an assignment within the portfolio.

All eighth grade students are responsible for presenting their completed portfolio before a panel which determines whether the student has met the "O'Farrell Standard." Panels include younger students who will one day be responsible for their own presentations. The student's advisor or "home-base" teacher is responsible for helping students prepare and rehearse portfolio presentations. Homebase teachers are encouraged to provide videos of successful presentations as resources to students as they prepare.

O'Farrell students use their portfolios to demonstrate what they have learned, but the structure of the portfolio and its implementation communicates to the students that their demonstration should be based on careful reflection and self-assessment. Students are provided access to exemplars of the work they are required to do and numerous opportunities to practice self-assessment skills.

When students have the skills, resources and opportunities necessary for quality self-assessment, it is hoped they will be more engaged by school work, more able and willing to take responsibility for their continued growth, and more motivated to continue their learning once they leave school. Chapter 4 suggests that this may be one of the more significant payoffs to continued implementation of alternative assessment strategies.

Challenges, Cautions, Caveats

Standards: The Missing Piece in Student Self-Assessment

Many CAC projects have given considerable thought to designing structures and providing opportunities for students to practice self-assessment. All too often, though, these structures and opportunities exist in the absence of the clearly articulated standards which make them meaningful. Students are asked to reflect on the quality of their effort without being assisted in understanding the characteristics of high quality work, or having been shown examples of it.

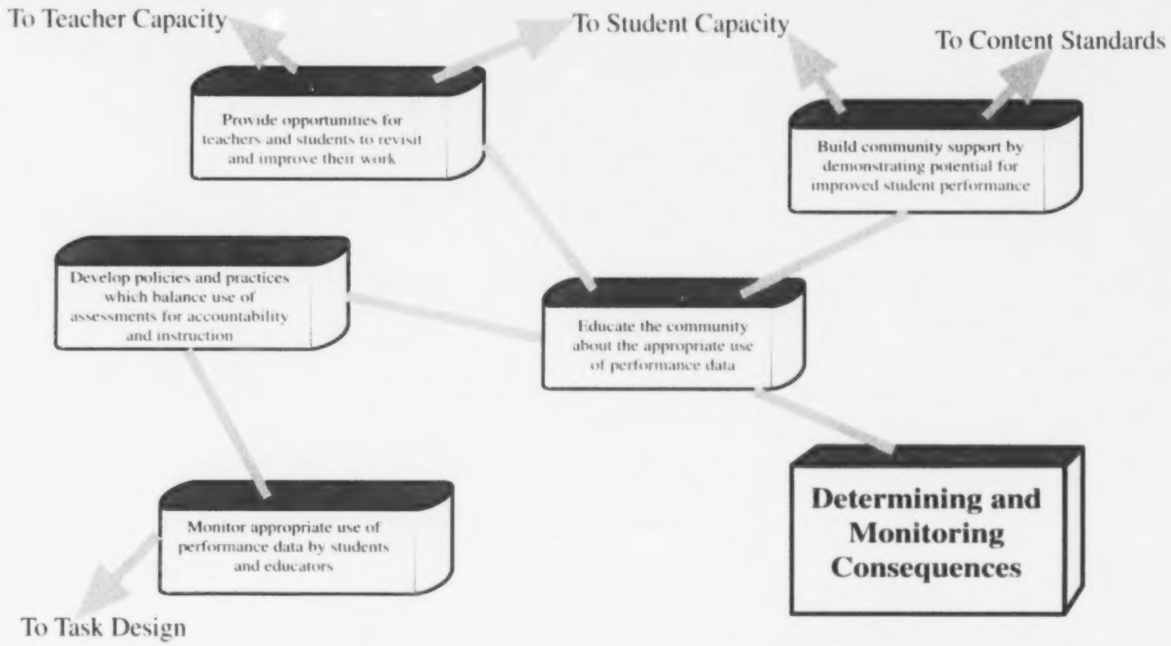
As mentioned in earlier sections of this report, teachers must be given opportunities to articulate and build consensus around standards of performance in order to help students know what is expected of them. Yet often this element of the work is given short shrift or ignored. While policy makers and educators struggle to determine what official standards should look like and who should determine what they are, teachers must be encouraged to provide exemplars and to engage students in articulating the characteristics of high quality work.

The Need to Support Teachers in New Roles

When students have the skill, motivation, and resources to assess their own work, the traditional relationship between students and teachers begins to change. Armed with knowledge about where they are and where they need to go, students can take much greater responsibility for their own learning and teachers are freed to serve as facilitators and coaches. Schools in which all students take major responsibility for their own learning will challenge teaching professionals to venture into new and perhaps initially uncomfortable roles. School leaders will need to consider ways to encourage and support the change process.

DETERMINING AND MONITORING CONSEQUENCES:

HOW CAN WE ASSURE THAT ASSESSMENT ENHANCES TEACHING AND LEARNING?


To Teacher Capacity
To Student Capacity
To Content Standards
Provide opportunities for teachers and students to revisit and improve their work
Build community support by demonstrating potential for improved student performance
Develop policies and practices which balance use of assessments for accountability and instruction
Educate the community about the appropriate use of performance data
Determining and Monitoring Consequences
Monitor appropriate use of performance data by students and educators
To Task Design

Many of the indictments against traditional assessments relate to the consequences of their use for individual students and the school program. The recent National Science Foundation study conducted in six urban school districts looked at both standardized tests used in district and state mandated testing programs and the textbook tests which students took frequently throughout the year. The study concluded that, contrary to the recommendations of curriculum experts, both types of tests "emphasize and mutually reinforce low level thinking and knowledge."

Additionally the tests were found to have an "extensive and pervasive influence" on instruction. Teachers reported spending a great deal of time on test-taking skills and instruction which was directly related to topics included on the tests. They also complained that these tests caused a narrowing or fragmenting of the curriculum, limited the nature of student thinking, or forced teachers to rush too much for students to learn well.

Perhaps of greatest concern was the finding that classrooms with large numbers of racially and culturally diverse students were affected significantly more by mandated standardized tests than other classrooms. The report concluded, "This suggests that the educational experience of students in high- minority classrooms is qualitatively different from that of students in low-minority classrooms." (Madaus, 1992)

Various other studies suggest that the large-scale, high stakes use of standardized tests contribute to both decreased motivation to perform and the sorting and labeling of students into "tracks" which deny educational opportunities to some students. (Darling-Hammond, 1992)

At their most benign, traditional assessments are thought to take valuable time away from instruction and provide little benefit to it. Yet standardized tests are generally the only source available to feed the public's seemingly insatiable demand for data on student performance. Test scores are sources for innumerable national and international reports of student achievement as well as headlines for the local paper. They are popularly believed to be the single most important indicator of the quality of a school's program. (Elam, Rose & Gallup, 1992)

School practitioners working on the development and implementation of assessment alternatives aspire to strategies which serve and actually enhance instruction. The San Diego City School District has drawn together the experience and priorities of numerous initial alternative assessment projects within the district to develop an "Assessment Vision and Policy" designed to guide the development of the district's assessment system. (Figure 25) It lays out the role of assessment in helping the district to restructure in order to become "more effective, productive, and accountable" and powerfully articulates the importance of assessments which contribute to improved teaching and learning. The vision and policy statement is a radical departure from traditional district assessment policies which stress the importance of assessment only for purposes of accountability.

If new assessment strategies are to have the positive consequences envisioned by this policy statement, schools will need to

- integrate teaching, learning, and assessment by providing opportunities for teachers and students to use assessment feedback to revisit and improve their work;
- make a commitment to an on-going monitoring of how assessment data is used by students and educators; and
- build community support for assessment systems designed to both report student progress and enhance teaching and learning.

In this section we will look at strategies used by CAC pilot projects to maximize the instructionally beneficial consequences of assessment.

Figure 25
San Diego City Schools - Assessment Vision and Policy
Draft dated August 4, 1993

Guiding Principles
The district believes that an assessment system should

- be responsive to all students served by the district; respectful of their multiplicity of experiences, backgrounds, and cultures;
- be dynamic and flexible, integrating teaching, learning, and assessment into one coherent system which demands high, consistent standards for all students;
- provide all students with multiple and varied opportunities to display their cumulative learnings through a variety of appropriate assessment techniques;
- prepare teachers, as part of a comprehensive district staff development program, for their active role in the development, scoring, and reporting of student assessments consistent with restructuring efforts designed to improve teaching and learning;
- inform students, staff, parents, community, and all other stakeholders about the purposes, forms, and appropriate uses of an inclusive system of assessment; and
- be aligned with and reflect the shared tenets of the California frameworks, California Learning Assessment System (CLAS) and the New Standards Project (NSP)

Assessment Policy
The Board of Education is committed to a district assessment system that is fair, objective, and accessible to all students. The assessment system shall provide all students with multiple and varied opportunities to demonstrate their knowledge and unique abilities. These assessments will link curriculum and instruction in ways that reflect the best instructional practice and document ongoing student progress. The system will monitor the effectiveness of educational programs, provide the educational community with information to improve teaching and learning and promote the success of all students.

BEST COPY AVAILABLE

Integrating Assessment and Instruction — Creating Opportunities to Use Assessment to Improve the Work of Teachers and Students

Cycling Assessment Data into Instructional Planning

Recognizing the need to go beyond accountability, many CAC pilot projects have included in the design of their assessments, or within their implementation strategies, structures for feeding assessment data back into the instructional program.

Developers of the GATE Portfolio in the Oceanside School District are drafting guidelines and procedures to encourage the use of the portfolio not only for selecting students for the GATE program but for planning each student's instructional program. The goal is to provide an individualized learning plan for each GATE student which will reflect knowledge of that student's strengths and needs as identified from portfolio data.

The interpretation of student performance in the ACE K-2 Assessment Project results in the identification of goals which can be used to guide future instruction. Teachers do not record a student score, but set goals for future observations, lessons, or experiences for students based on their analysis of student work. (See example of ACE K-2 Interpretation Guide, Figure 9) Implementation support workshops often include an activity entitled "Teaching While Thinking About the End". After interpreting several student performances, participants consider which instructional strategies were most helpful in preparing students to perform well and in which areas the curriculum or instructional program needs to be revised to assure that students improve. Participants then share successful strategies with colleagues and/or collaborate on the design of new activities.

Revisiting a Performance

Strategies for providing useful assessment data to teachers and students are enhanced by opportunities to revisit a previous performance. At O'Farrell Community School in San Diego, students have informal opportunities to rehearse their portfolio presentations with an advisor, and should they fail to meet the standards of the portfolio presentation, are given additional opportunities to present their work.

The ACE K-2 Project demonstrates a similar commitment to the opportunity to revisit a task after receiving feedback on an early attempt. Task prompts are three-tiered with each "Try" scaffolded to provide varying degrees of assistance and guidance from the teacher. Early prompts are designed to prepare students to perform later ones. Students are encouraged to consider the ways they might improve on initial attempts, and project plans include providing participating classrooms with exemplars of student work for this purpose. An example of a three-tiered prompt is shown in Figure 26.

Another mechanism for encouraging students to revisit and build from earlier work is to increase the number of points during the instructional/assessment process where students receive feedback. The Grossmont Coordinated Science Portfolio Project in its second year of development decided to compile and grade portfolios at least twice during the school year in order to assure students the opportunity to get feedback and improve their work before the final submission.

Monitoring How Assessment Data are Used by Students and Educators

It is probably true that many of the negative consequences of traditional assessments were neither intended nor anticipated. If alternative strategies are to avoid a similar fate, educators must take responsibility for monitoring the consequences of assessment as an on-going function of the assessment system. Although few CAC pilot projects at this writing are deeply enough embedded in instructional practice to determine all possible consequences, a few projects provide examples of how monitoring an assessment's use can influence its design and implementation.

After using portfolios for one semester, teachers in the Monte Vista High School Integrated Science Portfolio Project conducted a survey of student attitudes towards the use of the portfolio. The survey asked students to articulate if and how their participation in performance tasks and their responsibilities for compiling portfolios changed the way they interacted with people around them, with their environment, and with the learning process. The Mt. Diablo Unified School District conducted an informal study of the ways in which teachers used the data from the Curriculum-Based Math assessments in their classrooms. The survey inquired about the use of assessment results with students and the perceived value of individual student scores. A similar district-wide study

A Work in Progress

Figure 26
Excerpts from Scaffolded Performance Prompt
ACE K-2 Alternative Assessment Project

Growing Up Prompt — Try 1:

We have spent a long time talking and thinking about how we've changed as we've grown older. We've read lots of stories in which people grow and change. Today you are going to start a project to show what you've learned. You will write your very own book which tells about you and how you've changed since you were a baby. Your book should include a lot of ideas about the differences between how you look, act and feel now and how you looked, acted and felt when you were younger.

Each page in your book could say: "When I was little I... But now that I'm big I..." Your job would be to fill in the sentences with many different ideas. Or you might want to express your ideas in your own words. As you work on your book, you will want to talk to others to get help and ideas, but be sure that your book has at least some ideas of your very own.

When your book is finished, your teacher will ask you to read it to her or to your friends in class.

Prompt — Try 2:

We have spent a long time reading, talking and thinking about how people change and stay the same as they grow older. Now you can start your own project to demonstrate what you've learned. Your job will be to do a project that tells your own life story of growing up. Part of your project will tell about how you have grown since you were a baby and part will predict what you will be like as you grow older. You will want to think about the many ways you have changed, and might change, and in what ways you will stay the same. Think about how your looks, ideas, abilities and interests might change, and be sure to use a lot of these ideas in your project.

You can do any kind of project which you think will help you share what you have learned, but you must include some original writing that is an example of your best work. When your project is finished, your teacher will ask you to describe it to her or to share it with your friends in class.

Prompt — Try 3:

We have spent a long time reading, talking and thinking about how people change and stay the same as they grow older. Now you can start your own project to demonstrate what you've learned. You have already done a project which tells about how *you* have changed and will change. This time you will be asked to think not just about yourself, but to report on the life of someone else. You might choose to share the life story of a familiar character, or report on a real person — someone you know, or someone famous. Use your project to show the many ways this person changes or stays the same as they grow older. Think about their looks, their ideas, their abilities and interests and be sure to use lots of those ideas in your project.

You can do any kind of project which you think will help you share what you have learned, but you must include some original writing that is an example of your best work. When your project is finished, your teacher will ask you to describe it to her or to share it with your friends in class.

will be conducted in the Redwood City School District in proximity to administration of the Performance-Based Assessments in the Spring of 1993. This survey will ask teachers about the impact of the district's new scope and standards and performance assessments on their use of specific instructional practices in the classroom. These studies are designed to be used not only to assess the impact of the assessment on instructional practice but to identify areas where the assessment might be revised to better enhance teaching and learning.

Building Community Support for Comprehensive Assessment Systems

The survival, and therefore the real educational impact, of instructionally sound assessment depends on community support. Parents and policy makers must be convinced that instructionally sound assessments are valid and fair and worth the commitment of resources necessary to do them well.

BEST COPY AVAILABLE

Several CAC districts have begun work with school boards and community groups to demonstrate the potential of new assessment strategies to contribute to improved student performance. Teachers in the Language Arts and Science Performance Assessment Development Project in San Diego City Schools conduct Town Meetings to inform community members about performance assessments. In these meetings, parents engage in many of the same hands-on assessment activities required of their children in the classroom and have opportunities to see how rich data can be gathered from observing students at work on these kinds of tasks.

In order to build support for the continued development of the portfolio project at Ben Franklin Middle School in the San Francisco City School District, the school community has been invited to observe the potential of the project to help students reflect on and take responsibility for their own learning. Parents are asked to work with students to prepare the presentations of their portfolios. They review the student's work and provide feedback on its specific characteristics and strengths. After receiving this input, students present their portfolios to panels of adults which include representation from parent advisory groups, the district office administrative staff, and community members. During the exhibition, students use evidence from their portfolios to show how they have addressed pre-determined content standards. This portfolio exhibition format was field tested with a small number of students in the spring of 1992 and will be conducted with all eighth grade students in the spring of 1993.

Many CAC schools and districts have recognized the importance of working with parents and community leaders to secure support for assessment reform. In the absence of easily aggregated student performance data, these schools have chosen instead to bring to the attention of the public the range of valued student outcomes not addressed by traditional assessments and the quality of data available when students are challenged by performance assessments. School board or site council meetings designed for this purpose are often centered around the display and analysis of meaningful student work samples.

Judging from observation of several of these community forums, and the very number of districts which provide encouragement and fiscal support for continued development and experimentation, many parents and community leaders, at least for now, are intrigued by the potential of performance assessment to enhance student performance.

Challenges, Cautions, Caveats

"You're a '2', John. Have a nice summer."

The only way to assure that the consequences of the data from performance assessments will be used to improve teaching and learning is to *structure it* to do so, and to *be vigilant in monitoring* that it is actually used in that way. Decisions made at the point of structuring tasks and designing scoring and reporting frameworks have a strong influence on the ability of teachers and schools to cycle assessment data into instructional or program planning. Often decisions made to assure that a task can be "easily measured" or that performance data can be aggregated result in formats which inhibit the ability of teachers to use assessment data in classroom or program planning. While there seems to be no easy resolution of the dynamic tension between using performance data for accountability and for instructional improvement, efforts must be made to strike a balance between the two.

Likewise student use of assessment data to improve their learning must be structured and monitored. In projects where new assessment strategies were being field tested with students, it was not uncommon to see assessment data being used for little more than reporting. Throughout pilot projects there were only a few examples of efforts to cycle assessment data back into an opportunity for students to revisit and improve on a performance.

The Isolation of Many Assessment Development Projects

Unfortunately, many CAC assessment development efforts can be described as experiments which are isolated from other school improvement or reform efforts. Even in schools or districts involved in self-described "restructuring," the alternative assessment projects are often an off-shoot rather than an integral part of that effort.

The primary goal of school restructuring is the facilitation of meaningful student achievement for all students. School structures should be designed to assure that all students have ample opportunities to achieve desired outcomes. Oftentimes this means restructuring to build the capacity of teachers and students to make use of instructionally sound assessments or to structure consequences which enhance opportunities to learn. When assessment development efforts take place in isolation, projects risk being seen as "one more new thing" to burden already constrained and over-worked teachers.

CHAPTER 3

FIRST PERSON NARRATIVES

STORIES OF WORKS-IN-PROGRESS

FIRST PERSON NARRATIVES FROM SITE-BASED ASSESSMENT WORK: STORIES OF WORKS-IN-PROGRESS

The preceding chapter illustrates that assuring that performance assessments improve student performance means not only being mindful of standards, tasks, and scoring processes, but also structuring the consequences of assessment to build the capacity of teachers and students to improve their work.

Drawing on the experiences of CAC projects, these sections have presented both a conceptual map for building an instructionally sound assessment system and illustrations of some of the strategies used to address its various elements. The conceptual map will, we hope, offer schools involved in their own assessment projects a framework for self-evaluation and some strategies they might apply to their own efforts.

What the conceptual map makes no attempt to offer is a requisite point of entry for assessment development or a recommended course through its various elements. Although the collective work of the CAC's original pilot projects suggests that all of the elements are important, how and when any district or school addresses a given piece must depend on the purpose of the assessment project, the nature of work already accomplished, and the unique characteristics of the culture in which the project is embedded. The logical "next step" must be determined by participants as they assess the impact of their most recent effort.

The systemic nature of the conceptual map also suggests that the path through it is not linear, but recursive, and that work on each element influences the others. For example, the degree to which teachers have operationalized standards influences their ability to plan suitable instruction, to work with students to do effective self-assessment, and to score assessments reliably. Likewise, the design of scoring and reporting formats influences the degree to which assessment data can be used to influence the instructional program and the ways in which students can use the data in self-assessment.

Progress on a development project therefore, must be measured not in how many elements have been addressed, but by the degree to which the work on any element reflects or guides the best instructional practice and is owned and understood by those who must make use of it. Ignoring or giving little thought to the implications of standards will likely mean that either tasks will not assess what was intended or interpretations of student work will be unreliable. A lack of attention paid to the design of scoring frameworks can mean that teachers and students do not get the quality of information they need to improve their work. The projects with the greatest potential to improve student performance are those whose participants assess where work has gone astray, and return to the identified element, perhaps more than once, for revision.

The CAC's 22 first-year pilot projects started in different places, emphasized different dimensions, covered different elements, and pursued different courses through the map. The narratives written by project participants or leaders which follow describe the process of work in seven of the CAC's first-year projects. They were selected to represent the diversity of projects which come under the umbrella of the CAC and to give readers an opportunity to better understand the origins, progress, triumphs, and stumbling blocks in site-based assessment development and implementation.

'Discovering Teacher & Student Capacity" tells the story of a kindergarten teacher (Evergreen School District) whose work with the ACE K-2 Project began as an effort to find an assessment tool which would illuminate what teachers know about student achievement and has become a vehicle for expanding that knowledge and building her capacity to help students become self-directed learners.

"Creating Opportunities for Students to Learn and Practice Self-Assessment" is written by a teacher who worked along with colleagues in the Monte Vista High School Coordinated Science Program (Grossmont Union High School District), to develop a project-based portfolio to assess student achievement. The teacher begins skeptical that the integrated course can survive the consequences of traditional science assessments, and concludes that traditional assessments deny the students the opportunity to see their growth and progress and take responsibility for their own learning.

"A Place Where Dreams Happen" illustrates the role standards play in guiding school restructuring. O'Farrell Community School (San Diego City Schools) is continuously working to articulate what is expected of all students and how they will measure student progress toward the "O'Farrell Standard." The narrative illustrates how design of the school's curriculum and instructional program, as well as the student portfolios, are guided by decisions about what the community wants its students to know and be able to do.

In the report of work from the Chula Vista School District, "Developing Meaningful Tasks to Improve Learning," we see how an effort initially designed to prepare students for open-ended assessment items has begun to influence both the content of instruction and the school community's expectations for students. Similarly, in "Creating a District Culture of Collegial Inquiry" from the Hillsborough School District, we see examples of how assessment development efforts increase teacher collaborations, build a common understanding of standards for students, and build the capacity of teachers to take a greater responsibility for assessment.

The final two narratives describe comprehensive district efforts in assessment reform. One describes the "All Students Will Succeed" project in the Redwood City School District. District leaders in this socioeconomically diverse suburban community in the San Francisco Bay Area have committed numerous resources to a multidimensional effort to build a performance-based assessment system and support teachers as they strive to implement a congruent instructional program. The last narrative describes the work of the San Diego Unified School District. San Diego City Schools has both the opportunities and the challenges of integrating a number of national, state, and local assessment experiments in which they participate. The narrative describes what the district has learned from those diverse efforts and the primary challenges for the future of assessment reform.

These narrative voices illustrate the complexities of assessment reform but at the same time suggest the power of its impact on those engaged in the work. In these voices we hear both uncertainty and frustration, and the enthusiasm to continue.

Discovering Teacher and Student Capacity: The Story of the ACE K-2 Project

Cadwallader Elementary School is one of thirteen schools in the Evergreen School District in San Jose. Janice Goertz, the author of this narrative, teaches kindergarten at Cadwallader and served on the original development team of the ACE Consortium K-2 Alternative Assessment Project. Susan Hanna, Cadwallader's principal, has supported the school's continued participation in the ACE K-2 Project and provided leadership to schoolwide efforts in alternative assessment.

When Elyse entered my kindergarten class, she was a shy student who expressed her ideas primarily through her detailed and careful illustrations. She talked with no one except me, and never spoke to the entire class. She loved being read to and had an extensive home library, yet when asked to write her own books, she was reluctant to attempt anything beyond teacher-prepared, patterned books in which she drew illustrations and dictated the text. She was able to write her name and recognize some uppercase letters, but had no understanding of letter-sound relationship, and was very uncomfortable whenever I suggested she give it her "best guess."

By late spring Elyse had written not one but two books on friendship, the second after hearing the prompt for this performance. ("This one will have words. The other one is wordless.") She talked quietly with me and the other children as she worked, describing her work and what she was planning for her book. ("This is my fourth page. I have two more pages to go and I'll be finished. Then I make a title page.") Occasionally she interrupted her work to patiently help another child who was stuck. She also noticed mistakes in her own work and self-corrected. ("I messed up right there. I didn't leave a space between words.")

Elyse's book on friendship had a title ("MIY FRIENDS SHP BOOK BY ELYSE RIVAS"), front and back covers, and ten pages of illustrated text, which she wrote with her own temporary spelling and growing understanding of word boundaries. ["Afriend WILL HLLP UW SLVAPRIBLAMS" - *A friend will help you solve problems.*]

In assessing Elyse's work, I asked myself: Does Elyse approach language tasks with a positive attitude and a willingness to take risks? Does Elyse use strategies to make herself understood and to derive meaning from what others have to say? Does she have a sense of audience and approach language tasks as a problem solver? Does she use language effectively to express original thoughts? Does she use the conventions of language appropriately? These essential questions, developed by the ACE K-2 Alternative Assessment Project team, have made it much easier to observe and analyze both the process of Elyse's work and her final product. Keeping these questions in mind allowed me to structure the anecdotal records I kept of Elyse's work, and use them later when I assessed it. (See Figure 3, Chapter 2.)

Defining What We Value Enough to Assess

My teaching was not always so well informed or so consciously focused on the essential questions of what was important for my students to know. While I planned and presented an environment and a curriculum that were child-centered, my assessment program did not give me nearly as much information about each child as I needed. We did not give standardized tests. I observed children informally and used manipulative materials to test each child several times during the year on the isolated skills reported on our district's progress reports. The children's work reinforced the saying that "What you test is what you get." Our day was filled with literature, poetry, music and art, yet the children did not choose to write or see themselves as writers, express original thoughts, or approach language tasks with a positive attitude and a willingness to take risks.

So, when three and a half years ago, I sat in a room of 22 other K-2 teachers, each representing a different district, I was ready with a torrent of answers when the project leader asked, "What are the limitations of current assessment methods?" That question, and the group sharing which answered it, was revealing. I wasn't alone in my frustration, my care for my students, or my concerns about their progress. The next question we were asked solidified my commitment to this project: "What knowledge, skills, and habits of mind do you value that we do not currently assess?" We had been brought together as classroom teachers to develop an alternative assessment strategy or tool to use in the classroom, based on our experience and knowledge of children.

Thus began the most challenging, creative work of my professional life. We spent the next two meetings brainstorming what we thought was important for children to know and be able to do, and negotiating a consensus on these values. I realized much later that articulating these values is the key to the development of an assessment program which informs instructional decision-making.

The questions we asked in the values process simmered in the back of my mind as I worked with my students. My kid-watching became more focused on what was essential. I began to get anecdotal records which gave me, and the parents, useful information about the children.

Developing Tasks

Our initial work on values led to the next big question: If this is what we value, what will it look like and how will we assess it? We needed to develop some authentic tasks which reflected our values statements and which assessed a child's integrated use of language. We needed to create a test which we would be proud to teach to. As we wrote, we considered questions of authenticity and developmental appropriateness, as well as questions of standards and whether prompts were linear enough to get at cognitive changes over time. As we tried to balance all these questions and tasks, like a juggler trying to add just one more flaming torch to her act, we realized the enormity of the work we had been sent to do.

That fall we agreed to pilot one of the performances we had written in our classrooms, then bring samples of student work to our next meeting. I taught a unit on friendship, using literature, songs, and class brainstorming sessions, then, using the prompt, asked children to write a book about friendship. Using blank books prepared by me, and the sentence pattern "A friend is..." they dictated their words for me to write. I was not convinced that their books showed me their best work.

Doubts about my students' work, as well as my own, bothered me. One of the traits I had said I valued was risk-taking, but there was no evidence of this in our work. My colleagues came to our next meeting with similar concerns. Our first student work samples were unfocused, shoddy: we had to ask ourselves if our standards were unrealistic, or if we had failed to teach the students what they needed to know to do well.

At this point in the development process, I had two realizations: first, the prompt needed minor revision. We began a continuing process of reflection and rewriting. Second, we realized that our assessment tool had given us more feedback about our teaching than it had about our students. Before teaching another unit, we brainstormed everything we needed to teach to prepare students for their performances, always referring back to the essential questions developed in the values exercise.

Teaching While Thinking About the End

We returned to our classrooms to try again with the unit "Growing Up," which also required that each student write a book to complete the performance. I taught mini-lessons on the writing process, developing and using strategies for problem solving, speaking and writing with a sense of audience, research and data collection skills, as well as helping the children understand how they had changed since they were babies. I made a web containing all the necessary elements for successful performances, and put it on my clipboard with the essential questions to remind me what to look for as I wrote anecdotes about the children's work. I was both researcher and teacher, conferencing with individual students and teaching mini-lessons when my kid-watching showed they needed support.

As I reflected on our work in process and on the children's final products this time, I knew my earlier hunch had been correct. I had not over-estimated what children could do, but I had failed to teach them what they needed to know in order to complete the first performance. Their second performances were not perfect, but each student's work, including that of my ESL students, was of much higher quality.

Teacher as Researcher - Student as Author

The 1991-1992 school year was an opportunity to build on what I'd learned by sharing it with colleagues in kindergarten and first grade. Thanks to the enthusiastic support of our principal, Cadwallader was selected as a pilot site in the ACE Alternative Assessment Project, and I continued as part of a school team. Our team met to plan units, share what worked, and brainstorm solutions for what didn't work. Two regional meetings with teams from other pilot schools further inspired us.

Elyse was one of my students that year, a member of a class with several ESL students and children from wide-ranging backgrounds. We worked on "Growing Up" in the fall and "Friendship" in the spring. My teaching this time was much more focused on the presentation and risk-taking skills necessary for success with the performances; I expanded my preparatory lessons to include students writing class and individual books and performing choral poetry readings.

I had internalized the essential questions, so the anecdotes on my clipboard were more specific and more valuable records. My kidwatching during the first year had been focused on 6 to 8 students, and now I was getting anecdotes on all my students.

My students approached the tasks with varying degrees of enthusiasm and levels of experience, but something was different this time. They had begun to work more cooperatively, using each other as resources, not just relying on me for support. Elyse began to ask questions of me and the other children, especially during the mini-lessons.

In her conversations with me and the other children as she wrote her own books, she often made reference to other books we had read, occasionally bringing one to the table to find a word she needed. She had good visual memory, and could usually remember how to write a word, such as "friends" and "the" after seeing it several times. As I observed and worked with Elyse through the year, she began to use phonics cues as another tool in her writing, making it easier for others to read her books. She frequently wrote her own books, lists, and notes, and looked forward to sharing her work with the class and "whoever picks me up today." At the end of the year, she used her knowledge to write a book titled "HIW PEOPLE MAK BOOKS." Elyse has a growing portfolio which demonstrates that she is becoming an effective, creative user of language, able to express herself and to understand others.

Our principal has been wise in allowing us to become involved in authentic assessment at our own pace. While we continued the implementation of the K-2 Project, our upper grade colleagues worked with that original thought-provoking question, "What knowledge, skills, and habits of mind do you value enough to assess?" They began to understand the importance of articulating together what they want students to know, and then determined what a successful user of language looks like at each grade level. We all have become both better assessors and better teachers by working through this process. By May the upper grade teachers had developed a writing rubric at each grade level. The entire staff will implement portfolio assessment next year, continuing the process begun with the "values" discussion and enhancing our understanding of the progress of our students and our own instructional programs.

Creating Opportunities for Students to Learn and Practice Self-Assessment: The Monte Vista High School Science Portfolio Assessment Project

Monte Vista High School is one of ten schools in Grossmont Union High School District. Vivienne Deutsch, author of this narrative, teaches both traditional chemistry and coordinated science at Monte Vista, and is a leader in the development of alternative tools for assessing student work in coordinated science.

In the fall of 1990, I began teaching science to one class of ninth grade students as part of the Scope, Sequence, and Coordination's (S,S&C) 100 Schools Project. Two colleagues and I teamed to teach chemistry, physics, biology, and earth/physical science in a three year program based on solving common "real-life" problems. The first year of the program was exciting, successful, and popular with students and parents. We went from one class the first year to twelve classes the second year. Of course, with growth comes problems, and one we were faced with was determining an appropriate end-of-year assessment for our new coordinated science course.

Monitoring Consequences - The Need For Alternative Assessment

Each year students in all science classes in the Grossmont Union High School District are tested using an end-of-year assessment. The results of those tests are then made available to each teacher as an indicator of how well that teacher's class is doing on the specific concepts as compared to all other classes in the district. Using existing tests, we developed an end-of-year assessment for both the first and second years of the coordinated science program. An item by item statistical analysis was done comparing how well the coordinated science students did on the same questions as their counterparts in biology, chemistry, physical, and earth/physical science classes. There are some inherent problems with this comparison, however:

1. The traditional courses are all tracked. That is, all of the students in traditional science classes are in college preparatory or honors tracks. The coordinated science classes are non-tracked, made up of students with a wide variety of academic abilities.

2. The traditional classes are mostly 10th through 12th grade students. The chemistry and physics classes especially are usually composed of juniors and seniors. The coordinated science classes are composed of generally younger, less experienced students.

3. For most of the traditional science students, their current science class is not their first science course in high school, while for students in coordinated science, it may well be.

Evaluation had to be done, to provide hard evidence that coordinated science is successful. But the quality of the coordinated science program has to be evaluated on a different scale. This seemed to be the perfect opportunity to experiment with alternative assessment.

Identifying an Assessment Technology

Eighteen months after beginning the coordinated science courses, we added portfolios to the first year course. We used them primarily to collect all of a student's work. This alone was beneficial because the students develop pride of ownership in their portfolios and their own place in the room to keep their things.

Toward the end of the semester, we broadened the purpose of the portfolio. Students were asked to select what they considered to be their best laboratory write-up, writing sample, cooperative group work, math skill (math problems or a graph), and an independent selection (a piece of work that they were particularly proud of or enjoyed). We then asked students to evaluate all five of the pieces that they selected. These five assignments remain in the portfolio and accompany students to their next year's science class.

Using a broad base of resources to inform our work, we developed rubrics to evaluate the portfolios. We also developed a self-evaluation and a check-off sheet to assess the portfolios and determined that the portfolio would be the end-of-year assessment for coordinated science. (See Figure 23, Chapter 2.)

How Well Did This Work? Teachers and Students Use Assessment To Improve Instruction and Learning

After being involved with alternative assessment, I can see that what I previously considered to be quality performance in my traditional chemistry class may have been only a good memory or well-developed math skills. Now I can offer students the opportunity to present what they've learned in a different mode and check for true understanding of the concepts I teach.

I watched students blossom in the coordinated science class. Students who did poorly on traditional tests gained confidence and began showing evidence of understanding and insight in the subject area. Students who would probably not have survived the college prep chemistry class worked hard and improved their performances throughout the semester. They didn't quit!

Their self-evaluations were remarkable. All of the comments usually made by teachers or parents were made by students. We repeatedly read, "I should have been neater," "I should have researched this more," "I should have checked my spelling," and, "If only I had started on this project sooner." We are hoping that through the self-evaluation process students will become more self-aware and better motivated.

I believe the atmosphere in my classroom during finals week best summarizes the differences between traditional and alternative assessment. My traditional chemistry students left class glad that another hurdle on the way to college was over. They may have been discouraged that they did not or could not do better on a difficult multiple-choice final, a test that did not give them an opportunity to show what they had really learned. No wonder we were discouraged.

The coordinated science students, however, spent their last day reviewing their own work and realizing how much they had learned during the year. They revisited their successes and the things they liked the best. They evaluated these pieces and how they could improve their academic and interpersonal skills to achieve more success the next year. The coordinated science students left school this year with a sense of optimism and planning for next year. Their teacher did, too!

Where Do We Go Next?

We intend to start the portfolios sooner and to include a laboratory skills checklist. This year we will check off lab skills attained throughout the year as they are attained and tested. The documented skills will be recorded on the portfolio cover. Future science instructors will have a clear record of which skills have been attained and what work still needs to be done.

In order to maintain some content continuity we intend to align our curriculum with The Inquiry Performance Tests Assessment Inventory (IPTAI), developed at the Israel Science Teaching Center at Hebrew University in Jerusalem. We will develop a three-year plan for teaching the science processes identified in the IPTAI and present them in a spiral manner, with students completing the beginning levels before moving on to more advanced levels. Student performances will be reflected in their portfolios.

We also plan to explore students' attitudes toward science using attitudinal surveys. And building on the success we have seen in school sports programs, we will encourage goal setting as a tool for success. Using their portfolios to review their progress, students will set individual and class goals.

How Are We Accomplishing This Innovation?

In our third year of implementation of the coordinated science program and after our work with portfolio assessment, I feel light years away from the type of instruction and assessment I used for years. The choice to change our curriculum produced the impetus to explore exciting advances in science education.

I am extremely fortunate to work as part of a team of six people who are actively teaching coordinated science. Team teaching calls for tremendous amounts of conferencing and discussion. The result of all this collaboration is the evolution of new ideas. Support from our department chair, counselors, and site and district level administrators allows us to take risks and try new experiments and strategies.

Without full cooperation from the superintendent to the classroom teacher, this level of innovation would be incredibly difficult. We need high levels of communication within schools, with other professionals in our district, and beyond. We have found support everywhere we looked, and we feel an obligation in return to share the information and strategies we have found.

A Place Where Dreams Happen: O'Farrell Community School

Karen Volz Bachofer is a program evaluator in the Planning, Research, and Evaluation Division of San Diego City Schools. She has worked closely with the staff of O'Farrell Community School as a participant/observer from the inception of their work, and is the author of this narrative.

> *Dream with me if you will... Can you see a school that children want to go to, that teachers want to go to, that parents want to go to, where teachers work closely together to provide the best education possible for* *all* *kids, where students succeed academically, where adults and children work closely together, where people smile, where the feeling is warm and the environment is safe, where decisions made about children are shared decisions, where responsibility is shared, where creativity is high, where gender and race are just part of your uniqueness and not a barrier to your learning, where there is an extended family of teacher, family, community, agency, where classrooms are exciting places to be, where teachers are turned on to teaching, where students are the number one concern...?*
>
> Bob Stein
> O'Farrell Community School
> Chief Educational Officer

Structuring A Student-Centered Community School

O'Farrell Community School: Center for Advanced Academic Studies, which opened in September 1990, is the first middle level magnet school in San Diego City Schools whose focus and instructional program have been determined through the collaborative efforts of teachers, administrators, and community members. During the school's first year of operation, 1990-91, approximately 440 seventh grade students attended O'Farrell. During the 1991-92 academic year, O'Farrell's richly diverse student population consisted of approximately 1350 sixth-, seventh-, and eighth-grade students in the following distributions: 33 percent African American, 32 percent Filipino, 18 percent Hispanic, 9 percent White, 5 percent Indochinese, 2 percent Pacific Islander, and 1 percent Asian. More than half of O'Farrell's students qualify for Chapter 1 services and approximately 12 percent have been identified as limited English-proficient.

O'Farrell Community School is divided into nine educational families, and all but the most severely emotionally and physically handicapped students receive the same high-level, enriched curriculum within the family structure. A daily instructional block links the subject areas of Humanities (language arts and social studies), Technics (mathematics, science, and computers), and physical education into an interdisciplinary core curriculum. Industrial and fine arts, music, and foreign language (Spanish and Tagalog) share a Discovery block each day. These interdisciplinary programs are tied together through schoolwide eight-week thematic units, which include pride, choices, relationships, and rites of passage. All students have teacher-advocates and meet in homebase advisory groups on a daily basis, and all-school town meetings are held every other Friday morning to share information and strengthen school identity. Innovative scheduling allows teachers in every educational family to have daily common preparation time to develop interdisciplinary curricula, make educational family decisions, and meet with students and parents. Each educational family is free to adapt its daily schedule to facilitate the instructional program.

In order to achieve a reduced class size, O'Farrell has chosen not to fill the traditional vice principal and counselor positions at the school. Instead, funds which would have gone to these positions are used to supplement the school's staffing budget, reducing the student-teacher ratio during the core instructional block to approximately 23:1. As a result, the leadership and administrative duties that would traditionally be handled by a vice principal and head counselor are shared by the teaching staff. This includes all counseling and discipline services, which are provided by teachers within educational families. Each educational family has a teacher assigned to one of three leadership roles: team leader, curriculum leader, or social/emotional leader. The position that normally would be described as the principal is held by the "Chief Educational Officer (CEO)" who serves as the schoolwide "keeper of the dream," the leader for all of the educational families, the advocate for the students, and who fulfills schoolwide administrative duties. The school and the San Diego County Department of Social Services have formed a partnership that provides on-site access to social service agencies and direct assistance in addressing the social and emotional needs of middle-grade children and their families. These services are housed in a Family Support Services Wing on the O'Farrell campus.

The O'Farrell program features a student-centered, multiethnic learning environment, which provides for the social, intellectual, psychological, and physical needs of middle-level youth. O'Farrell staff are committed to ensuring that all students (including special education, gifted and talented, and limited English-speaking students) have the same access to and success in an academically enriched curriculum. Further, they guarantee that all students who complete the three-year course of study will qualify for enrollment in course sequences that lead to Advanced Placement courses in high school. Assessment of student learning focuses on attainment of educational outcomes and takes a variety of forms, including portfolios and exhibitions.

Assessing Student Progress

As the school began its second year of operation in September 1991 (and tripled in size!), the O'Farrell staff began to focus increased attention on the ways in which student progress was being measured. They were specifically concerned that the districtwide standardized testing results would not allow O'Farrell students to demonstrate the skills and knowledge gained during their years at the school. O'Farrell students, they reasoned, would have just spent two years in a learning environment that stressed thematic, interdisciplinary curriculum, hands-on experiences, portfolio work, exhibitions and projects, community service, and teamwork. How could progress in these areas be documented and shared with parents and the community? In addition, O'Farrell students were used to demonstrating their competence through a variety of assessments from performance assessment tasks to exit interviews. Would reliance on high-stakes standardized testing as the sole measure of progress send the message to students, parents, and community members that the alternative methods of assessment used at the site level were somehow inconsequential or had no meaning?

Because the school's first class of eighth-graders would soon be leaving O'Farrell, school staff were also concerned about the problems that might arise during the transition from middle school to high school. They worried that many of O'Farrell's students might not be assigned to advanced-level classes, especially if performance on the district's standardized tests was used by high school counselors to determine course placement. In order to ensure that capable students had access to advanced-level coursework in high school, O'Farrell staff decided that it was important to articulate a "standard" for achievement that would demonstrate students' readiness for such coursework.

Articulating Content Standards

And so, O'Farrell staff members spent the 1991-92 academic year identifying the skills, knowledge, and "habits of mind" that students must have by the end of the eighth grade that would allow them to be successful in challenging high school classes. Ideally, they thought, the "readiness" characteristics required to exit the middle school environment should match the characteristics needed for entrance into an advanced high school environment. O'Farrell staff has articulated these "requirements for success," which are collectively known as "The O'Farrell Standard." It is hoped that, at some point in the future, high schools will routinely use attainment of The O'Farrell Standard as the only requirement necessary for enrollment in college preparatory courses.

Preparing Students to Meet the O'Farrell Standard

A curriculum plan has been developed to ensure that students successfully achieve The O'Farrell Standard by the end of their three years at the school. This curriculum is built on the assumption that

- there is no hierarchy to the teaching of critical thinking or problem solving;
- a project-oriented curriculum allows for multiple entry points to learning;
- curriculum should be constructivist; that is, students should be actively engaged in constructing their own meaning rather than waiting for teachers to give them the right answer; and
- "Advanced Studies" at O'Farrell is based on deep understanding of content rather than vertical acceleration through courses.

The curriculum plan proposes that only four "subjects" be taught at O'Farrell. In order to reach The O'Farrell Standard, students must demonstrate competency in each of these areas by the end of the eighth grade. The areas include

- Self-Expression: the ability to express ideas and feelings in written form, supplemented by other media;
- Academic Reading: the demonstration of strategies to understand new and difficult written material;
- Quantitative Reasoning: the ability to apply data analysis and mathematical generalization to problems; and
- Research Report Writing: combining the above skills to produce a thoughtful research paper.

Students will practice the skills from all four areas during their three years at O'Farrell, and their performance and products in each area will be described at three levels: entry, intermediate, and advanced. Entry level performance describes the efforts of a student who is new to O'Farrell and is not familiar with the four basic skill areas. Students working at the intermediate level demonstrate some experience in the skill areas. Work conducted at the advanced level indicates readiness for college preparatory work at the high school level. It is assumed that most students will require three years to be able to perform at the advanced level, although the levels are not grade specific and are not intended for placement purposes. Rather, they are used to describe academic performance at a point in time.

Plans For The Future

Staff at O'Farrell view the development and implementation of The O'Farrell Standard as a dynamic and continuous process. At this point, they have identified the skills necessary to succeed in college preparatory courses in high school, developed The O'Farrell Standard, and conceptualized a general three-year curriculum plan organized around thematic units to assure that students meet the standard. However, there is still much work to be done and many questions to be answered.

O'Farrell's continued involvement in the PACE Portfolio Project and the New Standards Project, in addition to continued site-specific staff development, are contributing to the development of a schoolwide assessment system, O'Farrell's goal for the 1992-93 academic year.

Developing and Using Meaningful Tasks to Improve Learning

Chula Vista Elementary School District, in San Diego County, focused its alternative assessment efforts on preparing students for open-ended assessments in math and science. Janet Abbott, the math/science coordinator, writes about the impact of that work.

As part of a five-year plan to update its assessment program, Chula Vista Elementary School District became involved in the Greater San Diego Area Consortium of the California Assessment Collaborative. After deciding to focus our attention on mathematics and science, we formed committees of teachers to discuss possible assessment projects, and began the process of developing, field testing, and scoring open-ended mathematics problems for grades 3-5.

Developing Meaningful and Fair Assessments

Committees worked together to write math problems using the criteria presented in *A Sampler of Mathematics Assessment*, published by CAP and the State Department of Education in 1991: Problem situations must be interesting, have multiple entry points to meet individual student needs, have multiple solutions which allow and encourage students to create original responses, and must identify the audience for whom the students are writing.

Prewriting activities were prepared for generic open-ended problems in both math and science, and also for certain specific problems. To make sure that we knew how students should respond to the problems, and that students knew what was expected of their responses before they began to write, we drafted tentative scoring guidelines based on how we hoped students would respond. These guidelines helped us to refine our items. CAP's generalized rubric, a scale of 1 to 6, was used as a model.

We field tested the problems with students in several schools in our district. Responses were scored collaboratively, with judgements on the quality of student work based on the depth of mathematical thinking and understanding in students' writings and drawings, on creativity, and on students' abilities to solve problems in a variety of ways. (See Figure 17, Chapter 2.) We identified anchor papers for each of the six designated scores and used these to score other papers along with the CAP rubric.

Field testing the items was an extremely important step, which we repeated several times as we revised items based on our analysis of student responses. One problem in our first field test was "What does 3x5 = 15 mean to you?" Finding that this was not specific enough, and wanting to provide another entry point for students to respond to the problem, it was revised to read, "Your younger brother has just asked what 3x5=15 means to you. Explain to him what you think it means. You may draw a picture if you wish." When responding to this version, some students drew pictures and some didn't, making consistent scoring difficult. In the third field test, students were instructed to illustrate their explanation with a drawing.

We are continuing to revise the problems and scoring guidelines, and will use these to prepare rubrics for each problem. (See Figure 8, Chapter2.) We plan to expand the writing, administering, and scoring of open-ended problems to a larger group of teachers, by having staff members currently using the items train other teachers in their use.

Developing Student Capacity to Use Assessment to Improve Learning

Student outcomes are exciting. The problems help students become motivated and interested in their work. Students are learning to express themselves by writing about and illustrating their ideas in math and science. They are developing confidence in their ability to think and communicate about math and science and thus raising their own expectations in these subjects. Students take time to answer these problems, and are learning to assess their own responses. A "justification" box is included in many problems which encourages students to review and evaluate their thinking and writing. They are excited about the opportunity to demonstrate the in-depth understanding these problems demand.

We hope to further involve students in the assessment process by having them assist with setting standards for their work, scoring it, and revising their work to achieve the standards.

Teacher Analysis of Student Work Improves Instruction

Teachers also find using these assessments very rewarding. It is not until teachers actually use the problems, see student responses, and score them that the real insights occur regarding the need for instructional change. As teachers use the problems, they analyze the skills and concepts they need to teach so their students are successful. Teachers modify their instruction to match the goals and objectives being assessed by the problems. In particular, they shape their instruction to have students use more manipulatives and calculators, and to allow more time for students to share thoughts and work together as they respond to the problems. Teachers are observing their students at work to evaluate their level of understanding, and to help further refine their instruction. They now describe and evaluate student work on the basis of actual performances rather than on homework, quizzes, and tests. The district recommends that teachers share student work on open-ended problems during parent conferences.

Each teacher in the district is encouraged to administer and score at least one open-ended item in math and science during the year. We were able to provide the necessary staff development to an entirely new group of teachers when we asked them to administer the assessment during a summer school session.

Improved Student Performance Builds Community Support

Developing and using these assessments has had a significant impact on our district. Teachers, parents, and students are developing a new view of assessment as an ongoing, integral part of instruction. Much less emphasis is placed on standardized tests and much more on performance assessment. The quality of work produced helps to raise the expectations of both teachers and students, and the School Board, the press, and the public have been enthusiastic about the changes. We are in the process of altering the district's progress reports to reflect the changes in our assessment procedures. These changes started with kindergarten and are proceeding through the other grade levels.

This project has helped us become much more critical about what is important for students to know and be able to do, and how students demonstrate their knowledge. As we have gone through the process of developing and piloting open-ended performance tasks in mathematics and science at the K-6 level, we have examined what constitutes acceptable performance and what instruction is capable of producing it. We are delighted to see many teachers who hear about the process requesting to be involved in it. It is an enormous task to provide the staff development to make this possible, yet the teacher and student enthusiasm generated by this work has made its continuation a top priority in our district.

Building a District Culture of Collegial Inquiry

Hillsborough City School District serves students in kindergarten through eighth grade. Carl Zon, Assistant Superintendent and author of this narrative, has been instrumental in creating a climate for systemic change in his district and in building linkages with other districts.

The teachers and administrators in Hillsborough, a small school district located twenty-five miles south of San Francisco, have made a strong commitment to the changes called for in the California Curriculum Frameworks. For the past decade, our district's membership in the Assessment, Curriculum and Evaluation (ACE) Consortium has empowered us to plan and implement systemic change at each of our local school sites.

External and Internal Collaborations

The ACE Consortium has long provided a model for what schools can do when they think beyond the traditional structure of isolated local districts to the potential of collaborative effort. In the fall of 1989, members of the ACE Consortium Board established "the development of alternative assessments" as a goal. Given the significant paradigm shift from skills-based to whole language instruction, member districts planned a project to develop assessment strategies for English-Language Arts in the primary grades (now known as the ACE K-2 Project).

During the past two and one-half years, Hillsborough City School District has participated in the K-2 Project. As the Consortium's Project Coordinator, I have had the opportunity to work along with 22 gifted primary grade teachers representing the districts in the Consortium as they developed the performance-based assessment model. I have also been in a position to provide leadership for the dissemination and implementation of the project in Hillsborough schools.

We have learned that in order to help teachers implement strategies which require radical change, we must support them in many ways. This past year I named a teacher-coach to assist classroom teachers as they took on this challenging and exciting work. The teacher-coach met with the K-2 staff to plan the instruction they would need to provide for their students to perform well on the prompts. She assisted them in their classrooms with kidwatching, an essential component in analyzing student work, and she also aligned the curriculum demands of the K-2 Project's prompts with the district curriculum guides in History/Social Science. It is clear to me that the project has had a profound impact on teaching and learning in our district.

In another project, Hillsborough teachers worked together to develop an instrument for assessing critical thinking skills. We agreed on a set of attributes and behavior indicators which teachers then used to assess critical thinking skills while observing students solving problems in groups. As part of this work, teachers were also provided collaborative opportunities to look at videotapes of the students and discuss and assess what they saw.

As an offshoot of their involvement with each of these projects, our teachers have begun to evolve a stream of pedagogical consensus at a schoolwide level. In lieu of separate paths, teachers have begun to realize their strength in numbers and have begun to march to the beat of the same drummer. I believe the strength of their consensus has enhanced their self-efficacy.

Impact on Teachers and Students

Assessment has begun to drive instructional planning and curriculum design. I have watched our K-2 teachers select and develop instructional activities which have prepared their students for successful performances timed to occur as closure for major units of study. They recognize the pervasive impact and long term nature of their work and speak frequently of the need to provide support for students as they respond to increasing expectations. Our teachers have seen the need to devise manageable data collection systems, including portfolios, as their role has changed from 'sages on the stage' to facilitators and observers of learning. They are also considering change in internal and external reporting systems.

Finally, project participants have experienced the excitement and exhilaration of creating knowledge about learning and teaching. They have begun to assume control of assessment designed to inform instructional practice. More importantly, learners have begun to assess their own progress and clearly demonstrate the knowledge, skills, and attitudes which evidence their progress as confident and effective language users.

The depth and breadth of work which alternative assessment demands clearly exceed the resources of a small school district. But our participation as a member of a multi-district consortium has helped fill many gaps. Our linkage with teaching professionals in other districts has enabled us to begin to plan and implement change in instruction and assessment practices.

All Students Will Succeed: The Story of Redwood City School District

Redwood City School District serves kindergarten through eighth grade students in a socio-economically diverse community in the San Francisco Bay Area. Mary Lairon, Assistant Superintendent, and Betty Casey, Resource Teacher for Performance-Based Assessment, are leaders of the district's assessment project and the authors of this narrative.

The Redwood City School District has embarked upon a major project to improve student learning, using a comprehensive performance-based assessment system as a blueprint for the instructional program. The project involves over a hundred and fifty teachers who have developed standards for student performances, the assessment tasks, and aligned curriculum materials. All efforts related to school site planning, curriculum and instruction, and staff development are also aligned with this project. Every teacher and principal in the district has been involved in staff development on the new standards and performance-based assessments. The project officially began in the fall of 1990 and the actual performance-based assessments were given to students for the first time in the spring of 1993.

Background

The winter of 1990 was a critical time for evaluation and decision-making in the Redwood City School District. The superintendent announced his retirement, leaving a legacy of accomplishment. During his tenure, he had encouraged the adoption of a mission stating that all students will be successful and had presided over the development of a long range strategic plan. Many innovations, including a state-of-the-art Primary Education Center and intensive staff development in cooperative learning and other alternative teaching strategies had been implemented. A language arts and math scope and standards program, based on the new state frameworks, had been adopted. Direct services for students at risk had been provided.

In spite of these innovations, student achievement, as measured by CAP and traditional achievement tests, continued to be low. Although many of us were dissatisfied with these traditional assessments and felt that there was a lack of alignment with the newly adopted curriculum and instructional strategies, we shared an intense desire, as outlined in our mission statement, to ensure that all of our students would be successful in high school and beyond. Yet, we were forced to face the fact that nearly one-third of our students dropped out of high school every year.

In the spring of 1990, Dr. Ronald Crates became the new superintendent. He believes that our primary goal should be to improve student learning and achievement and that all staff should be accountable to achieve that goal. He challenged us to develop performance-based assessments aligned with the innovative standards and instructional strategies in the California curriculum frameworks.

Dr. Crates' watchwords "Assessment drives instruction" led to a district-wide movement known as "All Students Will Succeed," designed to produce continuous and significant progress towards ensuring that all of the students in the Redwood City School District would be successful in high school and beyond. Our primary efforts turned to the revision of standards for student performance and the development and implementation of district-wide performance-based assessments. However, staff development, curriculum development and selection, school site program planning and implementation, and bilingual and special education programs were also aligned with this goal.

Developing Meaningful and Fair Assessments

Our first efforts focused on teacher committees who reviewed our district Scope and Standards. Two committees, one in Language Arts and the other in Math, worked with the California Frameworks and the NCTM standards to align district standards with these documents. The committees also made two other major decisions:

- Students would be assessed at grades two, five, and seven in both Language Arts and Math.
- All assessments would be given both in English and Spanish in grades two and five to accommodate the growing number

of Spanish speaking students in the elementary grades.

In a suburban school district with some distinct ethnic and cultural boundaries, this decision signified an important "coming together" of "Eastside-Westside" schools. The entire district reaffirmed its commitment to teach toward the same exemplary levels of the standards.

Our teacher committees spent much time discussing the form our new assessments would take. We wanted to monitor district progress over time and to measure all our standards in a more real-world, performance-based mode. In an effort to further define issues of reliability and validity, our district entered into a partnership with CTB MacMillan/McGraw-Hill. We worked with our consultants from CTB to develop many task types—from simpler short answer items, to projects, investigations, and complex reading analysis tasks.

During the spring of 1991, our teacher committees, with training by CTB, developed prototypes of the tasks we envisioned. A core group of teachers developed task writing skills and learned about potential areas of difficulty. As the development process continued, the prototypes were valuable in subsequent staff development efforts and have remained at school sites as references for district teachers.

Building Teacher Capacity

It soon became clear that we would need many tasks and that all teachers needed to be prepared for performance-based assessment. The task of developing substantial tasks for the whole district and providing staff development seemed enormous, especially when our district faced budget cutbacks. Grants from the San Francisco and Stuart Foundations allowed us to hire three district performance-based assessment resource teachers and to begin our development year. During the 1991-92 school year we focused on three goals:

1. To provide three days of inservice in the principles of performance-based assessment and our Redwood City Project to every district teacher.
2. To pilot our prototypes, score them, and develop rubrics.
3. To develop the final tasks which would be used in spring 1993 to provide baseline data for our district.

In addition to the development of performance-based assessments and our systematic professional development plan, the district also implemented other strategies designed to support teachers in their efforts to increase student performance. Our resource teachers worked at school sites with entire staffs, individual staff members, and grade level groups to further align their curricula with current assessment practices, in preparation for the summative PBA's in 1993. Several presentations were made to parent groups about the philosophy of performance-based assessment, as well as the timeline of the district's plan. The district resource teachers also produced a monthly "Assessment Sampler" newsletter which served to keep everyone informed on the district project and to allow space for teachers to share their successful assessment practices with other teachers.

During the year, teachers made it clear that while they understood and affirmed our district standards and the concept of performance-based assessments, they felt the need for some specific curricular support. They wanted lessons and sources for lessons which "taught our standards" in a more active, engaging way for students than traditional textbooks. As a result of this request, almost forty teachers worked during the summer of 1992 on Curriculum Directories. Language Arts teams worked to structure lessons around a piece of literature appropriate to each grade level and to develop assessments which could be used formatively in the classroom. In math, lessons were gathered for each grade level and for each strand in the math framework, and assessments were developed around the standards for that strand and grade level. These directories have not now been edited, translated, and distributed to teachers.

Engaging Administrative Support

In addition, throughout the 1991-1992 school year, student learning and achievement was the number one topic of all the meetings of district administrators. They received training in standards, assessments and how to incorporate the standards and assessments into site planning. The superintendent worked with the District Management Team to ensure that site plans were developed with the primary goal of improving student learning and achievement. All principals and their school site councils revised at least two sections of their site plans to reflect the newly revised scope and standards and performance-based assessments. These two sections included newly designed student progress objectives which will serve as the basis for annual school accountability reports.

Challenges Ahead

The job still ahead of us is to pilot our tasks, refine them, and adequately train our 2nd, 5th, and 7th grade teachers for our first administration in spring of 1993. In addition, our district is faced with the continuing challenge of finding funding to carry out our assessment project.

Our district project is large in scope; while this presents somewhat greater difficulties than a school-based project, we believe it also affords us several benefits. It seeks to refocus the entire staff on a clearly defined set of standards and sets before all of us exemplary levels of student performance. Additionally, it defines good assessment practices which model good instructional practices. And finally, it provides a strong, locally-created and administered, districtwide picture of where we are, so we can plan for improvement in student learning.

Overview of Issues: A Superintendent's Perspective on Assessment Development

San Diego City Schools superintendent Thomas W. Payzant spoke at the UCLA Center for Restructuring Educational Standards and Student Testing (CRESST) conference on August 28, 1992 about the role of assessment in a large urban school district. His remarks have been edited for inclusion in this study. Dr. Payzant has recently been appointed assistant secretary for Elementary and Secondary Programs, U.S. Department of Education.

San Diego City Schools is the eighth largest urban school district in the nation. We'll have between 125,000 and 126,000 students this year. We're close to 30% Latino, 19% Asian, with large Indo-Chinese and Filipino populations, 16% African-American, 35% Anglo. We have about 14,000 GATE students, 14,000 Special Education students, over 60 different first languages, 31,000 Limited English Proficient students, and about 10,000 employees.

We're growing; we're diverse; we're culturally rich and an exciting place to live and work. And we are confronting, like everyone else in California, a period of scarcity which makes our challenges profound. We've been at this restructuring and reform business for four years now, and I'd like to give you a quick view of how I see the overall effort.

A large bureaucracy cannot be changed one piece at a time, or even three or four pieces at a time. System-wide change must be undertaken all at once. Restructuring, or change, or reform is a dynamic, continuous process. All schools, divisions, and offices not only must be involved in the beginning, but also must be kept involved in the ongoing change effort. The connections have to be made. We haven't done that well in school districts in the past.

These are some of the elements that we're working with: teaching, learning, staff development, governance and collaboration, assessment, parent involvement, accountability, and integration of social services. Another way to look at this is to combine our work into five categories which we're using for our strategic planning: curriculum and instruction, assessment and accountability, health and human services, the total quality management concept for high performing schools in a high performance school district, and the notion of how you deal with public engagement and support.

There's a lot going on in San Diego because of a lot of our people who are doing things across the country. The reason why San Diego is attractive, and I'm very pragmatic, is that we've got a student population that will meet the needs of just about every field tester that wants to field test anything in research and development in the United States of America. You can find it all in San Diego, and you can find it all in individual schools in some instances, where you've got the complete cross-section in terms of socio-economic, ethnic and cultural diversity right there.

We have some guiding principles as we look at assessment development in San Diego City Schools. They are an integral part of the overall district restructuring efforts. We're concerned about equity issues. If the equity issues are not addressed in a place like San Diego, they are not going to be addressed anywhere, and they have to be front and center.

We're really interested in that seamless web we hear so much about: the connections among the three major components of teaching, learning and assessment. The primacy of teachers is the key to this whole effort. All of our restructuring efforts better have an impact on improving teaching and learning or they are for naught. We also place an increased value on teacher judgment and consistent standards for all students.

Particular concerns emerge about equity as we look at beginning efforts and alternative assessment. We know we are going to have to use multiple approaches, a variety of strategies in teaching, find more than one right answer. For most of my career, I've been looking for the one best way, and I think I finally saw the light three or four years ago; There isn't one and probably never will be. We have to work without the traditional constraints of time and the quick fix, we have to work in collaborative groups, bring a variety of heritages and experiences to the work of those groups, and understand standards and expectations beforehand.

Some of the things that we're involved with are

1) The California Assessment Collaborative.

2) The next iteration of the California Assessment Program. With great trepidation, I accepted Bill Honig's invitation to chair the task group that put together the proposal for the new CAP in California, and I was only saved by people who knew a lot more of how to do it than I'll ever begin to know. And because of good people, I think we've got a good plan and now the test will be to see if we can deliver it and implement it as proposed.

3) We're doing some interesting things with the California Learning Record.

4) And we're also doing things with Dennie Wolf and the Harvard/PACE Middle School Portfolio Project. We're field testing that in San Diego.

5) We are also a test site for the New Standards Project .

6) We have an elementary assessment development project in language arts and science funded by the Stuart Foundation.

7) And we are involved in the pilot in social studies with the CRESST people at UCLA.

The message I want to give is that we're trying lots of different things. Let me offer eight observations about what we have learned.

First of all, assessment development and collaboration takes lots of time; it takes lots of energy, it's hard work and it's costly. But, before we get too hard on assessment and all that's going on, remember if you do curriculum development work, it's hard, it's costly, it takes time and lots of energy. If you really focus on teaching strategies and how kids learn, it's hard, it's time consuming, it's costly and it takes lots of energy. So why should we expect the assessment effort to be any different? If we look at that seamless web and look at the three together, then perhaps we can get some economies of energy and scale by doing the three simultaneously.

Second, assessment was once viewed by many teachers to be the bane of their existence. What we're beginning to see with our development work is a new kind of excitement because teachers see the connections. No longer is assessment pitted against teaching and learning and in conflict with it, but part of that overall combined effort to improve teaching and learning for all students. Assessment, when it is an integral part of learning and teaching, can be really innovative and a very, very positive experience for teachers and students. They get hooked.

Third, we're finding that collaboration, though it is hard work, is reinforcing. And when people learn to work together, it is a powerful kind of thing to come out of isolation and find that other adults and other students are thinking and worrying about some of the same kinds of things that you are as a teacher or principal. The kinds of things we talk about in the classrooms as being collaborative (cooperative learning, teams of kids working together in team-building) carry over when teachers and administrators and others model after that same behavior and see the connections. The development of new assessment really enables us to move away from the isolating activities that have so controlled and defined the way in which schools are designed and work.

Fourth, we have a lot to learn about new approaches to assessments. I've mentioned the equity implications. We have to deal with what the long term effects are on students, on teaching and learning. We have many, many unanswered questions. And what are the policy implications of all this?

Fifth, parents have to be involved. There is a great deal of suspicion, particularly in communities of color, about new approaches to assessment, because there is an achievement gap; we all know it's there. There has been some success in beginning to close that gap, demonstrated by measures such as Norm Referenced Standardized Achievement tests. And when we switch gears, people in communities that have not felt well-treated by public schools are understandably suspicious that we switched the ground rules, and it's going to be okay for some of the White and the Asian kids, but the Latino and African-American kids are going to get the short end of the stick again. We've got to involve parents; we've learned that.

Sixth, assessment issues don't exist in isolation from issues being addressed in our overall reform agenda. You can't talk about assessment in a restructuring school district without dealing with issues of governance, teaching, learning, accountability, parent involvement, and integrated social services. And we're learning that we have to deal with assessment in the context of the whole.

Seventh, it's really nice to see the researchers and practitioners, university people, people from research and development centers, teachers and principals working together in real schools with kids in the ways that we've been talking about for a long time. It's really beginning to happen. And it happens in a place like San Diego because you've got key people who can have a foot in both worlds, and get along with both kinds of people, and kind of shepherd it back and forth. And we have good people coming in from outside that understand schools as well as the world of the universities and the research world.

Eighth, there's still a lot of skepticism about what's going on. And we're sending mixed messages, and I'm part of the mixed message. We still have Chapter 1 schools, many of them in San Diego. What do we have to do for Chapter 1? We have to give grades 1 through 12 the Standardized Norm Referenced Achievement tests, and we pay attention to those results. We still use them a lot, and we still use them for purposes that we've traditionally used them for, at the same time we're intellectually saying we're wrong in the way we're using them. And then at the same time, we're pitching performance-based assessment and it takes a very, very sophisticated and trusting teacher to work his or her way through all of that mixed message and really stick with it. I'm trying to help with that, but I've got accountability pressures and the political pressures to deal with. We have cut down our system-wide norm referenced achievement testing in grades 5 through 7, and if we can do something about Chapter 1, then the messages will begin to be less mixed.

Last point. There is a danger in all of this that unless we really see the connections and make this a part of the overall systemic reform effort, we'll become fragmented once again. So with all of these efforts that we're doing in different places around the district, what we're trying to do is to help the people make connections. The principles that may be applied in New Standards or in the PACE project and the AB40 Collaborative are all the same in terms of being able to find new and better ways to help kids and adults deal with the issues of what kids know and what they can do in a way that reinforces learning and teaching and improves it for all of our kids.

We're excited about where we are, and know we have a long way to go, but if we don't try it and work hard at it in a wonderfully rich and diverse community like San Diego, we won't get it done. If it can work in San Diego, it can work anywhere, and I think it will.

CHAPTER 4

ANALYZING BENEFITS AND COSTS

ANALYZING BENEFITS AND COSTS

Over the course of the last ten years, California's 1000 school districts and 200,000 teachers have made a significant investment in dollars, time and energy toward school reform. The curriculum frameworks and school level task force reports are the backbone of state level initiatives related to professional development, teacher preparation, administrative training, text and materials adoptions, school restructuring, and assessment reform. Districts and schools have organized in-service and community education programs, and struggled to provide the resources necessary for students to have access to thinking, meaning-centered curricula and instruction. Individual teachers throughout the state have taken up the challenge to learn new instructional strategies and in many cases have sacrificed personal resources in order to provide students access to materials and experiences which support a curriculum of richness and depth.

The real size of the investment to date is unknown. But as reform initiatives continue, inevitable questions arise about the relative costs and benefits of the various dimensions of the effort. In times of dwindling resources, policy makers and school leaders must look at assessment reform efforts in terms of the advantages and disadvantages of that work and consider what priority and level of support it should receive.

At the time of this report, there are likely to be very few educators in California who are unaware of the performance assessment movement. In May of 1993, the first stage of the new California Learning Assessment System (CLAS) brought performance tasks to fourth, eighth and tenth graders throughout the state. Yet in many ways the development of performance assessments is still in its infancy. Educators at the district and school level no doubt have many questions about how to find resources for such a complex undertaking at a time when basic programs are being cut. Even those who find arguments for performance assessments compelling have questions about

- how the increasing availability of performance assessments should influence the level or breadth of standardized testing conducted in their schools;
- the relative trade-off between involving local teachers in assessment development projects and purchasing commercially-produced performance assessments; and/or
- the depth and breadth of local development efforts; choices between funding small experiments using teacher experts, or more expansive efforts which expose large numbers of students and teachers to the innovation.

As one aspect of its work, the California Assessment Collaborative (CAC) has attempted to collect preliminary data on the costs and benefits of performance assessment strategies and to contrast that with what is known about the costs and benefits of more traditional assessments. Our data related to alternative strategies were collected in the CAC's initial 22 pilot projects. Project leaders provided categories and estimates of expenditures for a single year of work on their performance assessment projects. Surveys and interviews of project participants, as well as observations of the on-going work in these projects, provided opportunities to identify additional potential costs and benefits. (Note: Appendix A provides a detailed description of the CAC's data collection procedures.)

Performance assessment development can be looked upon as the task of creating packages of standards, items, and scoring processes or the *challenge of developing a system of interdependent dimensions which includes structuring the consequences of assessment to build the capacity of teachers and students to improve their work.* The difference between the two is that only the second assures improved student performance. When performance assessment development takes place in schools, there is a unique and critical opportunity to look at the costs and benefits not only of developing a performance assessment product, but *an instructionally sound assessment system.*

But it is difficult to estimate the eventual costs of a complex system from isolated works-in-progress. CAC pilot projects give us as much information about the cost of developing and implementing an instructionally sound assessment system as the cost of

BEST COPY AVAILABLE

building the Wright Brothers' plane told us about the eventual costs of the modern space program. Likewise, our description of benefits is constrained by our inability to yet see performance assessments in widespread use. Nonetheless, this initial development work highlights the complex dimensions necessary to an accurate analysis of costs and benefits and suggests necessary areas for further research. What follows is a framework for analysis of those important issues drawn from the early experience of a wide variety of assessment experiments.

No formal studies are required to demonstrate that the costs of developing and implementing instructionally sound assessment are high. We have drawn on the work of CAC projects to determine in what ways the benefits of their work for students, teachers and the school community might off-set some costs.

The Benefits and Costs of Traditional Assessments

American children are assessed in many ways. Most tests are devised by teachers to see how well their students learned material recently taught. Quizzes, take-homes, oral presentations, chapter or end-of-book tests, and final exams punctuate the school calendar.

Studies suggest that teachers rely heavily on their own assessments for student achievement data (Shulman, 1980) and that those assessments include observations, teacher-made tests, and tests included with curriculum materials. (Stiggins,1992; Madaus, 1992).

Teacher-made tests most often make use of short-answer, matching, multiple choice and true/false items. Less than 1% of all test items reviewed in a study of 400 teacher developed tests (Fleming and Chambers, 1983) required essay responses. In the same study, 80% of teacher-made test items were found to assess knowledge of terms, facts, rules and principles. Few questions required students to demonstrate application of that knowledge.

A National Science Foundation study of classroom math and science assessments showed that, depending on grade level, 53% to 87% of math teachers reported using textbook or teacher-made tests at least once a month. Textbook tests were found to be used slightly less by teachers of science, and teacher-made tests in science were used in equal proportion to those used in math. Teacher-made tests were found to be similar to textbook tests in their emphasis on low-level thinking skills and recall of information. (Madaus, 1992)

In addition to classroom assessments, students are generally required to take standardized tests which originate outside the classroom and are used to guide decisions such as student eligibility for special resources or qualification for admission to special school programs. Most standardized tests use multiple choice items. Standardized test scores are often aggregated to describe the performance of classrooms, schools, districts, or states and have come to play a major role in shaping public attitudes about the quality of American schools and the capabilities of individual students.

California's students take standardized tests as part of the state assessment program and again as part of local district assessment efforts. District assessment programs usually make use of commercially-produced, machine-scored, multiple-choice tests. In some districts these tests are given in all major content areas at all grade levels, in others only in a few grades. Some districts use them only in those areas required to determine proficiency in basic skills or to identify participants and evaluate the progress of Chapter 1 or other externally funded programs.

In a survey of 43 districts in San Diego County, 95% of the districts used commercial tests. The number of grade levels tested ranged from two to twelve. All districts which used commercially produced tests gave the tests in reading, written language, and mathematics. Fifty-six percent of the districts gave the entire test battery which usually included reading, written language, mathematics, social studies, and science.

Benefits of Traditional Strategies

Multiple-choice standardized tests and short answer classroom assessments offer several advantages:

- Questions can be answered quickly, allowing tests to **cover a broad range of topics within a content area.**
- Multiple-choice tests yield scores that can be **efficiently reported as a single number or percentile** which represents the students' overall performance on that set of tests
- The tests yield individual **scores which facilitate comparisons** with other groups of students.
- Multiple-choice tests are thought to be relatively **inexpensive to administer.**
- Multiple-choice items are constructed to represent unambiguously right and wrong responses and therefore can be **scored using inexpensive technologies.**

- Commercially developed standardized tests **assure the validity of correct responses** through extensive review and pilot testing.

- Efficiently reported scores help **reveal the unequal distribution of educational resources and opportunities.**

Costs Related to Traditional Strategies

District Testing Programs

Because of the variance across districts in the number of grades and subject areas tested, the cost of standardized testing is difficult to describe. A recent study conducted by the Office of Technology Assessment (OTA) of the US. Congress looked at the costs associated with standardized testing in one large urban school district during the 1990-91 school year. The study detailed the following costs:

Administration Materials and Services — Outlays for contracted materials and services of test publishers, supplies, and time for district personnel to coordinate test administration and study and report results. OTA estimated these costs at $12 per pupil in the district studied.

Administration Time — The portion of total teacher salaries attributable to time spent administering tests. OTA estimated these costs at $13 per pupil.

Test Preparations — Cost of time spent by teachers in test preparation, estimated at anywhere from 0 to 3 weeks for each administration. These costs were estimated at nearly $100 per pupil .

Opportunity Costs — Opportunity costs refer to the estimated value of the activities which are displaced by time involved in testing. In the case of standardized testing the estimate of opportunity costs would depend on the degree to which preparation for and administration of the test were seen to contribute to the instructional program. Opportunity costs in the district testing program studied by OTA were estimated to be between $13 and $110 per pupil depending on teacher perceptions of the degree to which test preparation activities were relevant to course work.

In addition to the costs identified in the OTA study, additional costs related to district use of commercially-produced standardized testing might include:

Teacher Orientation — In some schools and districts teachers are provided brief orientations to test administration and management. These sessions, which are often conducted at the school site, cover issues of management of test booklets, scheduling of testing periods and make-up tests, and distribution of materials and supplies. The costs associated with these orientation sessions are the portion of faculty and administrative salaries attributable to orientation time and the related opportunity costs.

Test Security — District policies and publishers' guidelines require attention to assuring that test booklets and student answer sheets are kept secure between testing periods. Some policies require these materials to be kept in locked facilities and re-distributed each day of the testing period. The highest costs related to test security are incurred in those situations where policy requires the presence of test proctors, individuals assigned to watch test administration and distribution of materials to prevent cheating.

The California Assessment Program

All categories of costs described above apply to the student assessment program in California. In 1972, in order to reduce both the actual and opportunity costs of state testing , the California Department of Education developed a matrix-sample format. In matrix-sampling, no one student takes all items of the test. School level scores are derived by combining the scores of all students who take the test. Matrix sampling minimizes disruption to the instructional program, while providing valid data for purposes of program evaluation.

Classroom Assessment

Test Writing — Because of the idiosyncratic nature of the efforts, no estimates are available of the time involved for teachers in writing tests. Some teachers re-use tests made from one year to the next, while others design items to reflect the focus of work done by individual classes. The costs associated with test writing are the proportion of teacher salaries attributable to writing time.

Administration/Scoring — Judging from the research described above, teachers generally choose test formats which can be administered and scored quickly and reliably. Text-based tests usually have answer keys available which can be used to facilitate scoring. Many teacher-made tests are designed to be scorable on Scantron machines which read and mark multiple-choice answer sheets in seconds. The availability of the Scantron has made it possible to use lower cost clerical support staff to score tests. Potential costs associated with administration and scoring include salaries attributable to staff time to administer and score tests and purchase and upkeep of scoring equipment.

Purchase of text-based assessments — Many texts include tests at the conclusion of chapters or units of study. In addition, many publishers have developed placement tests, skill tests, and end-of-book or unit tests which can be purchased as supplements to a textbook series. These assessment materials are generally congruent with material presented in the text, and therefore to pass, students must have the text in order to prepare. When course grades depend on these text-related assessments, it might be said that costs should be estimated to include both the purchase of assessments and the textbook required to pass.

Opportunity Costs — Opportunity costs are the value of those activities displaced by classroom testing. Because classroom teachers generally write or select classroom assessments to align with their instructional program, it would seem that opportunity costs for classroom assessments would be low. But, as we have learned, teacher made and text-based classroom tests generally assess only low level skills and recall. Therefore when the curriculum aims for higher-order skills, the opportunity costs of these tests may be significant.

Delayed Costs of Traditional Assessments

Previous parts of this report suggest costs related to traditional assessments which cannot be measured in dollars. Traditional assessments have been found to influence instruction in such a way that many students are denied access to opportunities to practice higher-order thinking skills and engage in meaningful instructional tasks. (Madaus, 1992) Standardized tests are used to sort some students into low "tracks" from which school drop-out rates are high. (Darling-Hammond, 1991)

Although it is difficult to assign dollar amounts to these costs while students are still in school, it is not hard to imagine where the account is eventually rectified. Business and industry report spending increasing sums to train workers in problem solving and teamwork. Colleges spend more each year on remedial programs. And society pays for welfare and social programs for those unable to find work.

Both text-based and multiple-choice standardized tests are as common in schools today as textbooks, paper, and pencils. Although the indictments against their use are not new, the increasing demand for information about the achievement of individual students and schools, and the lack of commonly accepted alternatives, has meant that the perceived advantages of using traditional assessments have prevailed over the costs.

A Framework for Analysis of the Benefits of Performance Assessments

Proponents of performance assessments believe they hold great promise for guiding school reform and improving teaching and learning. Although it is too early to know whether CAC projects will indeed have such impact, through their work we have opportunities to see specific examples of the ways in which school-based assessment projects might contribute to the implementation of an instructionally sound assessment system.

The comments and observations of individuals do not tell much about the actual impact of this work. To understand the scope of beneficial impact, further research is required in each project to determine whether a good number of teachers and students were similarly affected and the degree to which the perceived benefits could actually be attributed to a change in assessment practice. The framework provided below provides an interesting contrast to the costs and challenges of performance assessment and may help structure continuing research.

Enhancing Opportunities to Learn

CAC projects offer examples of the potential of site-based assessment projects to enhance opportunities to learn by increasing the use of performance-based instructional strategies, expanding instructional opportunities for special needs students, increasing teacher sensitivity to diversity and bias, and raising teacher expectations for student performance.

Increased Use of Performance-based Instruction — Many project participants report that as a result of their work with performance assessments, they have increased their use of instructional strategies designed to teach or help students practice using prior knowledge to construct meaning from new material and effectively communicate what they know to others.

> *I seek out projects like Street Law and Project Business that align with the California framework and our portfolio project. . . . Now I'm combining immigration with the constitution. I'm seeking out primary source documents that give a more complete understanding of issues.*

In task development and scoring sessions CAC project participants often shared instructional strategies which they found effective in helping students prepare for new performance assessments. At least two projects built into their work processes for compiling, duplicating, and distributing congruent performance-based instructional ideas contributed by project participants.

Increased Instructional Opportunities for Special Needs Students — According to a survey of CAC project teacher-participants, nearly two-thirds felt that their involvement in the alternative assessment project had a positive impact on students with special needs. Projects which were specifically designed to explore assessment issues for special needs students, as well as those in which special program teachers worked closely with colleagues from the regular program, provided examples of the ways in which performance assessment can provide a vehicle for both assessing, and then addressing, the quality of the instructional program for special needs students.

In one portfolio project, after reviewing student portfolios to assess the kinds of tasks given to special education students, the faculty noticed that work done in Special Education generally focused on isolated skill development and was not aligned with the performance-based instruction going on in several classrooms.

The portfolio itself provided a vehicle for staff to address this issue. Using the work kept by the student in the portfolio and the criteria for success which accompany the performance tasks, the Special Education teacher works with students to help them succeed at the same kinds of complex tasks required of all students.

In another project designed to explore assessments which help illuminate growth in second language acquisition, participants discovered by reviewing student work in portfolios that English learners were often denied access to teaching strategies specifically designed to meet their needs because they were "reclassified" into the regular education program too soon. Teachers in this project spoke of the consequences of some of their work with portfolios:

> *Redesignation has slowed down and students are feeling more successful. An overall benefit to all students is that with alternative assessments, students have knowledge of requirements and buy-in of results.*

> *We are now doing strategic planning for LEP. We have added more sheltered English classes and have more administrative backing. Our SIP proposals finance sheltered English workshops and curriculum development for teachers. We are also working on a welcome center for assessing new students. There is also greater awareness of the needs of LEP among the staff through presentations at faculty meetings.*

Increased Sensitivity to Student Diversity — Many CAC projects were initiated because participants were sensitive to the fact that traditional assessments often ignored the diversity of language, culture, and learning styles students brought to school, and were therefore biased. The work in CAC projects often enhanced and broadened these sensitivities.

Observations provided numerous examples of participants arguing for revisions to assessments in order to eliminate some bias or expand opportunities for more students to succeed. In one assessment of mathematics, some teachers argued (as yet, unsuccessfully) for the inclusion of oral directions for open-ended math items so that students with weak literacy skills could still demonstrate their knowledge of mathematics. In another project participants pushed for inclusion of more opportunities for students to use graphics and other displays to demonstrate knowledge of US History.

Even false starts provided lessons about diversity. In one project designed to assess oral language proficiency, teachers devised and field-tested a task which required students to hold conversations in a second language with student partners whose names had recently been drawn from a hat. Review of video tapes from the field tests revealed that the task was unfair to students from cultures in which it was unusual or even inappropriate to hold a gregarious conversation with a mere acquaintance. In later revisions of the task, teacher participants considered having students choose their own conversation partners.

Increased expectations for student performance — The relationship between teacher expectation and student performance is well known. When teachers believe students can perform well, they have a better chance of doing so. CAC projects provide examples of the ways in which participation in performance assessment projects can expand teacher expectations for student performance.

Scoring sessions provide teachers opportunities to review student work samples from classes other than their own. After analyzing numerous exemplary samples, teachers have access to graphic evidence that students are able to perform at a high level and must confront their own responsibility to provide opportunities for all students to do so.

> *When I looked at this prompt, I said to myself 'Come-on get real, my kids can't do that.' (But) The scoring session was really revealing. It seemed I was right about some kids, but we really saw some great work. It made you want to find out what those teachers were doing!*

In a project which encouraged teachers to use portfolios to identify gifted and talented students, one teacher remarked:

> *There's a joke at my school that I think my whole class is GATE. . . . I mean I know we need to know who they are, but at the same time, to have this kind of assessment and curriculum makes you see the exceptionality in every child. There's something strong in every child, and it may not be academic and it may not be art, but they can do something well, and I think for our at-risk students that's so important. I think it causes you to look at every child by doing this (the portfolio). We have to.*

Enhancing Teachers' Content Knowledge and Expertise

CAC projects also provide examples of the potential of site-based assessment projects to enhance teachers' content knowledge and instructional expertise by helping teachers articulate standards and by providing a catalyst for professional growth.

Improved Ability to be Explicit About Standards — When criteria for a strong performance are explicitly stated, they provide teachers with guides by which to plan instruction. Participation in collegial assessment development assisted many teachers to use more precise language to describe the characteristics of high quality work. As teachers participated in rubric development, criteria for "well organized writing" came to be stated more precisely as "states problem and follows with supporting evidence," "has distinct beginning and closure," "uses headings and subheadings to label ideas." "Solves problems creatively" became "did not rely solely on models provided by others" or "considered more than one possible solution." These more precise statements are not only more helpful to students struggling to improve their work, but better teacher guides for instructional planning.

Alignment of Professional Development with Student Performance — Performance assessment development projects sometimes provided a catalyst for teachers to pursue professional development. As the result of opportunities to look closely at student work or to consider expanded content standards, teachers engaged in self-assessment and recognized areas in which they needed to improve their own skills.

> *Teaching sixth grade has made me realize what I don't know about science. Elementary teachers don't major in science. It forced me to do my own inservice, my own reading and I take classes at SDSU. Now I've found that the more you know the less you know. This project has put a focus on my professional growth. I am preparing myself for my students.*

As an outgrowth of one district-wide assessment project teachers are encouraged to select professional development opportunities specifically aligned to areas in which students demonstrate weak performance.

In addition, other pilot projects built in opportunities to develop subject matter and pedagogical knowledge as part of the assessment development process. In order to strengthen teachers' ability to make effective instructional use of a rubric for narrative writing, one project focused a good part of its development work on building teacher expertise in the characteristics of children's literature. In another project, teachers reported that the informal sharing and discussions at portfolio development sessions increased their knowledge about second language acquisition and appropriate instructional strategies to support language growth

The Creation of "Cultures of Collegial Inquiry"

The experience of participants in CAC projects suggests that site-based assessment projects have the potential to ameliorate much of the isolation teachers experience in schools and create cultures in which critical analysis of student work is both valued and expected. The potential for building these "cultures of inquiry" comes from assessment projects which provide vehicles for building consensus on standards, create structures for collegial work, encourage conscious instructional decision-making, support the role of teacher as facilitator, and increase teachers' knowledge about assessment.

Increased Teacher Consensus on Standards — Few teachers get opportunities to learn whether their standards for student performance match those of their colleagues. Students describe teachers' standards as disparate, labeling some "easy" and others "hard graders." Repeatedly teacher participants reported that the assessment development projects provided valuable opportunities to negotiate standards for student performance with professional colleagues. These exercises not only seemed to help build trust between teachers, but provided opportunities for individuals to critically examine their own work.

> *(The project) made me think more about what I am teaching. Do kids really need to know this? Am I covering little facts or am I giving the big picture?*

Teachers from the development team in one project were asked to plan orientations for colleagues new to the assessment strategy. Asked what parts of their original development work were critical experiences for the new teachers, participants unanimously felt that all teachers needed to participate in an exercise to determine consensus on what students should know and be able to do. A teacher from this project commented

> *The benefits (of this project) are in everybody knowing what you value for students clearly before you get started . . . to clearly define where you want students to go.*

Creating Structures for Collegial Work — Performance assessment development projects often challenged the traditional isolation of teachers. Teachers came together for the purpose of negotiating standards, designing and scoring tasks, or learning assessment skills and sharing congruent instructional strategies. These collegial opportunities were chances to learn from and about one another:

> *The project has fostered peer coaching. Teachers are feeling more comfortable to ask each other, "Would you come help me?" "Would you come teach?" "Could I watch you?" That's a new experience for our staff. They feel comfortable with another teacher helping them. . . . I see that as tremendous personal growth.*

> *I have learned that high school teachers are people too! Working with teachers from K-12 has been great. . . . It's been great to have this group to communicate and share ideas. Because of it I have done a lot of communicating with other staff at my site about how to assess ESL students.*

Teachers from many assessment development projects recognized the need to reject the idiosyncratic nature of grading and interpretation of student work.

> *The most valuable part (of this project) is the opportunity to sit down with a colleague and student work and to think about it, and argue about its quality. The scoring showed we could get agreement on quality.*

Encourages Conscious Instructional Decision-Making — Some performance assessment strategies developed by teachers in CAC projects already provide rich data about what students know and need to learn.

> *When you ask students to write about their learning you find out about your teaching — you assess your own performance instantly. You get it right away if your students have got the concept.*

Armed with this kind of data, teachers in many projects reported using it to plan instruction to meet individual students' needs.

> *The most powerful benefit to any teacher is experiencing the link between instruction and assessment. This allows continual self-analysis of teaching methods/materials in order to show student growth. That analysis leads to more conscious instructional decision-making and in turn to empowerment of teachers.*

> *We recently had the opportunity to use the information contained in student GATE portfolios to hatch a new program for GATE students. Given the information in the folder the site team brainstormed instructional activities that we thought would benefit these students in their careers. The quality of the information available to these teams enabled them to generate a plethora of ideas that were then shaped into specific instructional activities, goals, and objectives.*

> *After giving the Fall assessment we got together and set priorities for student instruction. We used the assessment results to target students who would normally have slipped through the cracks. We had the extra knowledge and data to proceed with working with individual students. It made me realize some students who were doing better than I thought. I have now set higher expectations.*

> *I've had to wrestle with what exactly do I teach and how it can be measured. It takes an analysis of the goals of my teaching and planning activities to support those goals. I've found I need to use all kinds of approaches and I'm more scientific about deciding which to use. I tailor instruction to fit students, use different materials to meet different needs, and I have different tools to assess each.*

Support for Teacher as Facilitator — Many teachers who used performance assessment strategies in the classroom reported that the student assessment tasks guided them to organize their own work in new ways. Because the tasks challenged students to take much of the responsibility for planning and carrying out their own work, teachers were freed to become facilitators and observers of student learning.

> *It feels good to me as a teacher, my role is different, being aware of what resources are available and how students use them and each other, it's not all on my shoulders. We aren't telling them what to do, they can rely on each other.*

> *My curriculum is different from last year and I like it better. Students do more research and the topics are more student driven. I still give them guidelines of what they need to cover but students have a say in planning. They get excited about it because it is flexible. I used to be more teacher directed.*

> *It makes me less and less a controller of the class. I give kids experience at finding solutions, and walk away from those teaching days excited about the ways they solved problems.*

Increased Knowledge about Assessment — Participation in assessment development work was credited by some teachers for having helped to increase their general knowledge about assessment, its various uses and audiences. This knowledge contributes to their ability to be critical consumers of assessment instruments and data.

> *I see alternative assessment- as it's practiced in our school - as an addition to traditional assessment tools - not a replacement. There are some skills, like arithmetic and grammar whose gross*

> *level can be determined by traditional "quick and dirty" fill-in-the-blank type assessments. I think it would be highly unprofessional of the educational community to throw out these traditional assessments totally in favor of only the alternative assessments.*

> *In terms of content tested, many alternative assessments are limited - however, in the full range of thinking and problem-solving skills, alternative assessments give you a good picture of student progress. Overall, alternative assessment has tremendous possibilities for helping and enhancing instruction at the classroom level. However, aggregating to the district or state level is costly and often eliminates the detail and richness of examples that are the main potential benefit of alternative assessment.*

Improving Communications with Parents

CAC projects provided several examples of the potential of performance assessment strategies to improve home-school communications by providing better quality data on student performance and increasing opportunities to engage parents in supporting student learning at home.

Improved Quality of Information about Student Performance — In a few projects where data from new assessment strategies were reported to parents, teachers reported having a better quality of information to give. Instead of reporting only grades, these teachers shared examples of student work from portfolios, records of teacher observations, and examples of student responses to performance tasks. Several teachers commented that this kind of information better illuminated growth over time and was particularly reassuring to parents of students with emergent literacy or early English language acquisition skills.

Encouraging Home Support of Student Learning — Several teachers also noted that the quality of data available to parents from new assessment strategies facilitated collaborations between parents and teachers to improve student performance.

> *We gave the fall assessment just before parent conferences, we presented the results and (parents) could actually understand them. It (helped me make) prescriptions for parents and me to help students improve.*

> *The first conferencing sessions did much to enhance the home-school team and reinforced the parent as "first teacher." These were gratifying sessions providing opportunities for exchanging information, setting goals, and clarifying purpose.*

Guiding School Restructuring and Change

CAC projects provide examples of the potential of site-based assessment projects to facilitate change by both sustaining and motivating restructuring projects.

Helps Sustain Restructuring — Participants from several different projects described performance assessments as a key element in sustaining their school's restructuring effort. Having made a commitment to pursue integrated or interdisciplinary curriculum, teachers in several schools considered portfolios and complex performance tasks as opportunities to demonstrate what students could accomplish in these curricula, and to assess for themselves the progress of student achievement.

Writing about why she wanted to try using portfolios in an integrated science course one teacher explained:

> *I had seen science innovations come, go and mutate, and I wondered if the SS&C coordinated science would be a credible science revolution. I have found myself being skeptical, not of the merits of the new methods of science instruction, nor the need for a change in emphasis in science curriculum. But I am skeptical that the new approach can buck the already established system of standardized achievement tests.*

At least two schools which had restructured to form "families" or small sub-sections of students, all of whom report to the same group of faculty members, saw the portfolio as a valuable vehicle for facilitating communications between teachers within the family group.

Motivates Restructuring — A few schools within at least one CAC project provide evidence of the potential of alternative forms of assessment to give direction to school restructuring efforts. Recognizing that the traditional 180 day school year did not provide enough time for some students to achieve performance standards, and that students could benefit from having access to numerous exemplars of performance, three schools within one project organized non-graded, multi-age classrooms. The assessment tool helps to measure a student's progress within that classroom while allowing students to progress at their own pace. Less experienced students have the advantage of seeing exemplars of work produced by those more experienced.

Enhancing Student Achievement

The primary purpose and ultimate measure of the success of assessment reform is improvement in the ability of all students to do meaningful work. As action-research projects still in their infancy, few CAC projects yet provide opportunities to assess the impact on students' ability to reach clearly articulated performance standards. Nonetheless, CAC projects provide some evidence of the potential of their work to enhance student performance by increasing student engagement in learning activities, skills of self-assessment, and willingness to take responsibility for their learning.

Increased Engagement In and Motivation Toward Learning — Eighty percent of teacher-participants surveyed felt that students would be more motivated toward their work as the result of participation in performance assessments. Although it is unlikely that changes in student behavior can be attributed solely to the use of an assessment tool, it is significant to note the specific types of changes teachers observed.

One teacher's comment suggests that the project-like classroom assessment tasks provided models for students to continue what they considered interesting work on their own:

> *My students are excited about their writing and reading - and their parents make comments like "He's always writing books now at home. I don't know why!" They choose to write during free choice time. They read spontaneously whatever they see in their environment. And they support and encourage each other. This has to be the result of an assessment tool which values them as learners, which is just a part of the curriculum, which values risk-taking and problem solving and cooperative learning. In short, I see my students approaching tasks with confidence, beginning to learn to assess and value their own work (and that of others), and to see themselves as readers and writers. And they are in kindergarten!*

A teacher from another project suggested that the potential of performance assessments to improve student performance came through engaging students enough to persist through difficult tasks:

> *I watched students blossom in this class. Students that initially did very poorly on traditional forms of objective testing gained confidence and began showing evidence that they did understand the concepts and showed insight into the subject area. Students who would not have survived a (traditional) class were working hard and improving themselves even at the end of the semester. They did not quit!*

This teacher's comments seemed to be validated by some of her students who wrote about what they liked about completing a portfolio collection of performance tasks:

> *. . . being able to make a project of our good work. Being able to keep it for others to see because so often our work goes unoticed (sic).*

> *. . . it gave me a sense of completion. I didn't think I could finish the project because it was so long. Now that I did it, I am happy I did. I learned a lot about me and the work I can do.*

Increasing Ability to Self Assess and Take Responsibility for Learning — Among project participants surveyed, over seventy percent felt that use of alternative assessment strategies would increase the ability of their students to assess their own work. Several projects provided specific evidence of the potential of performance assessment strategies to help students assess their own work up against content or performance standards:

> *I have devised my own rubrics suited for the activities in my class. Students score each others' work, and their own. (They) give reasons for the score, suggestions for improvement and decide if they want to continue to work on it.*

> *Students are more involved. They understand what is assessed and know what they need to do and make choices to do it. They understand implications, they challenge themselves to write a few more words. They compare progress to themselves, not to others.*

Finally, several teachers attributed an increasing willingness on the part of some students to take responsibility for their own learning to their use of performance assessment strategies. Armed with the criteria for success and skills to self-assess, students were thought to be better able to identify areas in which they needed more work:

> *I have watched students' confidence grow. Students feel the ability to revisit their work and grow from their mistakes. They learn that evaluation in itself is a tool for learning to improve (not a label). . . . Students show full ownership and pride with their finished work.*

> *[An important benefit of the project is] . . . the buy in by students that comes from choosing the work to be assessed and having to self evaluate it. . . . Students grow in decision-making abilities and are empowered by their participation in the assessment. . . . Students evaluate their own progress, and are not comparing their work to others.*

One teacher's reflections on the contrast between students on the last day of school in her traditional college-prep chemistry class and those in the two-year, coordinated-science class, where they are required to complete performance tasks and self-evaluative portfolios, probably points best to the potential of performance assessments to benefit teaching and learning:

> *My chemistry students left my class being glad another hurdle on their way to college admissions was over. They will probably not take chemistry again unless required to do so. They may even have left discouraged that they did not, or could not, do better on a very difficult multiple-choice final. I felt disheartened that, after working very hard all year, once again my students did not do better on their final. I never gave them an opportunity to show me what they had really learned. They left not realizing the depth or breadth of their knowledge. No wonder we become discouraged.*
>
> *The coordinated science students spent their last day reviewing their own work and realizing how much they had learned during the year. They revisited their successes and the things they liked the best. They evaluated these pieces and how they could improve on them next year. They evaluated their own behavior in groups and planned on improving their interpersonal skills for achieving more success next year. The coordinated science students left school this year with a sense of optimism and planning for next year. Their teacher did too — it was a nice change!*

Thinking About Benefits

It is important again to note that many of the positive pay-offs to site-based assessment work were not experienced by all projects, nor is it true that all teachers or students in any one project experienced it in the same way. In order to use this analysis of potential benefits to further their work, and the work of other schools and districts who wish to learn from them, it will be important for the CAC to explore both the degree to which any potential benefit was experienced in a given project and what specific design or implementation characteristics made them possible.

It also should not go unnoticed that nearly all the potential benefits described above are attributed not solely to the *use* of performance assessments, but to the *participation* of students and teachers in site-based assessment projects. Much of the perceived positive impact came from the act of creating, debating, interpreting, and scoring with professional colleagues. That finding, along with the previously described challenges of assessment development in schools, points again to the need to see assessment reform not as an effort to develop and disseminate products, but a challenge to re-invent both what teachers do, and the schools in which they do it.

A Framework for Analysis of the Costs of Performance Assessments

According to project action plans, most CAC projects drew from a wide range of funding sources, including district general and categorical programs, special projects, and grants to support the development of their alternative assessment strategies. Personnel costs accounted for the major portion of expenses in all projects.

Personnel for Task Design and Development — Projects reported their largest expenditures for teacher stipends or substitute time used to free teachers to: design tasks, portfolio templates, and scoring schemes; and/or make revisions following informal classroom tryouts or formal field tests. Substitute teachers cost between $70 to $100 per day and teacher stipends ranged from $14 to $20 per hour. The number of teachers involved in projects varied. Some projects chose to limit their work to a small core development team while others included a large percentage of staff in hopes of expediting the implementation process. In most cases development teams included ten or more teacher participants. Estimates of involvement time varied from a low of three to a high of more than a dozen days per teacher per year. Even within a single development team teacher time varied, with some members contributing more days and hours than others.

Observations of work in pilot projects suggest that there will always be a tension between the actual and opportunity costs of teacher time and the requirements of good development work. Some projects would have benefitted from providing a good deal more time for teachers to articulate and build a consensus on content standards. As the result of insufficient development time, tasks often proved to assess skills other than those intended.

The cost of teacher time also often inhibited the number of revisions permitted before a task was finalized. In projects which supported numerous revisions, the question of burn-out arose:

> *Time is a big challenge. Its been a changing process, an evolution. Each of our questions and decisions begat more questions and it really mushroomed. At first it looked like we had a neat package, now its much messier, but better. We keep working at it and hope to keep from getting burnt out.*

Teacher Orientation — In addition to expenses for teacher time related to development work, some projects provided time for all teachers in the district, or at least those teaching at grade levels to be assessed, to participate in orientation activities designed to prepare them to teach-to or administer the assessment. Teachers were provided with prototype assessments to use as curriculum materials and other instructional strategies to prepare students for the upcoming assessment pilot. One district-level project dedicated four teacher work days to this purpose. Another held after-school sessions on a single day.

Teacher reactions to receiving alternative assessment strategies they had no hand in developing and/or no opportunity to preview suggest that all projects may require orientation efforts for teachers new to the assessment:

> *After giving the [performance-based] test, I have nothing good to say about it. It left the students very frustrated, even angry. It covered material they had not seen. I tried not to help them, but it was very difficult. I felt that it was a waste of my time and the students' time.*

Personnel Time for Administration, Interpretation and Scoring — Participants in many CAC projects talk about blurring the lines between assessment and instruction, suggesting that assessment activities might be so integrated into the instructional day and so congruent with instructional goals that it would be hard to distinguish distinct amounts of time for administration. Yet few if any projects have reached this goal.

In projects experimenting with curriculum-embedded assessments or portfolios, teachers wonder how to integrate assessment activities into their professional day:

> *The greatest challenge is the time factor involved in conferencing individually with children. If a school is going through restructuring, this would be a key point to address — How to manage time for each student on an individual basis in a classroom of 30+ students and only one adult.*
>
> *It is not as much the planning and preparation of activities that make an alternative assessment a*

challenge, but rather the one-to-one time necessary in execution and follow-up.

Projects which featured on-demand, open-ended items generally took time for analysis and scoring during after-school hours. One district-level project reported spending approximately $2.00 per student on substitute costs and stipends to give teachers time to develop a rubric, score, and record the performance on an annual assessment. A two-day scoring event involved over 80 teachers and several clerks and data entry personnel.

Any concentrated analysis of student work for the purposes of assessment takes time. In many projects that time was integrated within the participants' instructional or professional development day and was not estimated. A complete analysis of the costs and benefits of alternative assessment strategies will require further inquiry into the actual time and relative pay-offs to teachers like those quoted above.

Project Facilitators — In addition to classroom teacher release time, many projects supported district or school level project facilitators. These project facilitators were often teachers on special assignment or curriculum and instruction specialists. They managed the development projects, planned and conducted assessment development sessions and teacher inservice programs, supported teachers who were trying assessments out in classrooms, and consulted with external experts. In some projects these facilitators were hired exclusively to support the assessment project, but in most cases they carried out a variety of additional responsibilities in their districts.

Projects without designated facilitators generally progressed slower than those with the advantage of that assistance. Facilitators often took responsibility for consolidating the work of development teams, typing it up and getting it back to participants between development sessions. For teachers doing development work following a full teaching day, this was welcomed support.

Clerical Support — Project facilitators and teacher participants required clerical support to type, make revisions to, copy, and distribute the work of assessment development teams. In many projects, clerical support was provided as an in-kind contribution from district or other special program staff.

Consultants — Finally, several projects indicated expenses related to the use of consultants. Commercial test developers, university researchers, or curriculum or assessment consultants were contracted by several projects to provide additional expertise on topics which ranged from technical issues of assessment design to related subject matter and instructional strategies. These experts varied in their level of involvement, with some making occasional visits to observe and advise, and others providing on-going leadership and direction. Occasionally these consulting services were contributed by external agencies.

Actual dollar amounts for expenditures for personnel and consultant services varied in proportion to the scope of the project. One project developed solely by a small number of local teachers spent about $46 per student a year on teacher release time to develop, field test, and score math and language arts tasks. In contrast, another district, which involved large numbers of teachers in development, provided comprehensive orientation for all teachers, and worked with an external consultant, estimated expenses at nearly $108 per student.

Reporting Achievement Data — At the time of this writing few CAC projects had established procedures for reporting student achievement data to various audiences. Nonetheless, the portion of salaried time for teachers to report to parents or school or district administrators to report to community groups must be calculated in order to get an accurate picture of the costs related to implementation of alternative assessments.

Materials for Assessment Development, Administration and Reporting — All projects incurred expenses to reproduce assessment tasks and teacher training materials. Again, actual costs varied by project size, numbers and types of tasks developed, and the number of revisions made. Other expenditures were related to the production of assessment materials and the storage of student work. As teachers struggled to manage increasing amounts of data, they experimented with the use of various expandable files, storage crates, and boxes for filing and organizing student work.

While few projects considered their development work finished during the CAC's initial funding period, some administered their assessments to large numbers of students and collected and reported student performance data. These projects provided opportunities to look at the nature of expenses related to materials for administering new assessment strategies.

The costs for producing assessment forms, test booklets, administration manuals, and/or teacher support guides varied with project size and complexity. Production costs for three test booklets used in one district were estimated at less than $0.30 per student. The same project supplies teachers with individual student results and a grade level report that includes scoring rubrics for each item and exemplars of student work. The per copy cost of the report was approximately $1.75. Another smaller project spent approximately $1.60 per pupil to produce assessment administration manuals and student reporting forms.

Supplies Used in Assessment or Preparations — Some assessments required students to use instructional materials such as literature selections, math manipulatives, calculators, scissors, markers, or chart paper. These materials were sometimes supplied to students as part of test administration and sometimes assumed to be available in the classroom. Teachers reported that they did not always have these materials and sometimes spent their own money to acquire them. Neglecting to supply teachers with materials essential to the assessment dampened the enthusiasm of some teachers for the alternative strategies:

> *I, and all of us who are willing to change must be helped, supported . . . I will not spend any more of my own money to make changes the district and state want me to make!*

Because the need for such supplies varied even within a single project, cost estimates for these materials were not available.

Challenges which Imply Additional Costs

In addition to the categories of expenditures detailed above, an accurate account of the costs of developing and implementing instructionally sound assessments must acknowledge the cautions and challenges revealed in early development work. Each of these challenges suggest additional potential costs related to assuring that assessments enhance teaching and learning.

Addressing Inequities in Opportunities to Learn

Inequalities in opportunities to learn are brought to the surface by site and district based assessment development efforts. Teachers on development teams bring vastly different expertise and resources to their work. Assessments created by one group mirror rich meaning-centered curriculums, while others reveal that students are being taught only isolated skills. Student work from some classrooms provides evidence that students had access to a variety of materials, equipment, media and other resources, while others did not. Student self-assessment is structured, taught, and practiced in some assessment strategies and ignored by others.

Teacher participants in development work recognize that what students are expected to learn and the opportunities and resources they are given to learn it vary widely even within a single school, and that those inequities impact students' ability to perform well:

> *Children in classrooms that did a lot of writing were not threatened by the test. Students who are used to writing in math enjoy writing what they know. They like thinking up an answer and writing their solutions* ***if*** *they have done similar activities in the classroom.*
>
> *I believe performance assessment is an important practice and will benefit students. However, these assessments are more readily accepted by teachers who run student-centered classrooms. Our biggest challenge is to help teachers make the shift from teacher to student-directed learning and implement instructional practices advocated in the frameworks. Then these assessments will be seen by teachers as valuable and vital parts of the curriculum they teach.*

Under current school structures, even in schools with relatively rich resources, teachers struggled to create opportunities for students to practice the kind of skills promoted by new assessment strategies:

> *The greatest challenge is setting up an environment that allows time for the important aspects of portfolio assessment to occur. This means a time for students to generate their products, reflect upon what they have done, and engage in uninterrupted periods in which to process the tasks. Also time for the teachers to meet with each individual student to assist in the setting of goals and to celebrate the accomplishments that have been made.*

A look across numerous alternative assessment development projects shows that exacerbating the inequities in opportunities to learn is vast confusion over the meaning and role of standards. Many teachers assume that colleagues within a given project share common expectations for students, while assessment development or scoring strategies reveal that they do not. Others admit their belief that standards cannot and need not be shared. The vast majority, even within alternative assessment development projects, had little opportunity to articulate their own standards, let alone work toward consensus with others.

In calculating the costs of an instructionally sound assessment system which is fair to all students, it is necessary to consider costs related to articulating standards and assuring equitable opportunities for all students to reach those standards.

Building Teacher Content Knowledge and Instructional Expertise

Observations of the work in CAC projects suggest that not all teachers possess the curriculum content knowledge or expertise necessary to develop meaningful instruction and assessment tasks. Exercises aimed at helping teachers articulate content standards often reveal that critical elements are missing from their understanding of the discipline. Teacher-developed tasks sometimes emphasize recall of isolated facts at the expense of essential understandings.

Teacher participants in CAC projects were often aware of this shortcoming:

> *Most middle school math teachers are math majors or have trained extra/concentrated in math. I, as an elementary teacher, have five main subjects and five or six minor subjects on which to focus and train. I am very willing to learn but I am far from a mathematician. . . . I want to stay with the times, so to speak, but I need the support of the state and my district!*

Teacher participants also recognized the mismatch between traditional instructional practice, which often emphasizes a single correct answer, and the open-ended nature of alternative assessments:

> *I just got two big manila envelopes in the spring containing the tests! . . . I have not been visited, helped, or instructed in the assessment project. I just lucked out and got to go to the one multiplication workshop that gave me clues. I find it peculiar that for two years I have been given alternative tests to give but have not been helped with or exposed to the curriculum on which they are based. Isn't this the cart before the horse?*

An accounting of the costs of instructionally sound assessment must address the potential costs related to increasing teachers' knowledge of curriculum content.

Overcoming Cultures of Isolation

Teachers in most schools have few opportunities to collaborate on standards or on interpretations of student work. A teacher's professional day leaves at best 50 minutes for preparation and nearly no time to meet with colleagues. Traditional class loads require that teachers be familiar with the work of between 30 and 150 students. Although as described above, alternative assessment development efforts often provide some release time for collaborative work, in order to assure that assessments enhance teaching and learning, time for assessment must be woven throughout the professional day:

> *At my school we have all heard the word alternative assessment — given the choice we opted for alternative assessment over standardized assessments at first (everyone hates standardized tests), but then we realized that there was no time to do the alternative assessments — not just to do it, but a structured thing with time to think and develop activities. Therefore we switched — there was just not enough time to think and plan.*

Instructionally sound assessment requires that teachers have opportunities to participate in assessment activities which are embedded within the professional day and year, to analyze student work in depth, and to plan instruction based on the data:

> *Teachers are moving away from being disseminators of information to facilitators of learning. The most challenging, and most rewarding problem has been trying to restructure the classroom to allow for learning situations that will allow for the observation of student growth and development.*

Schools committed to using assessment to improve teaching and learning communicate its importance by structuring and supporting collaborations among teachers, holding teachers accountable for the analysis of student work up against standards, and reducing the number of students whose work teachers must get to know. Teachers commented on the challenge of creating school cultures embodying these expectations without the support of school leaders:

> *[The greatest challenge is] our supervisors: If the administration doesn't know and understand the importance of implementing alternative assessment, then it's hard to convince other colleagues to do it.*

Possible costs associated with building school cultures which support instructionally sound assessment include providing collaborative teacher time, reducing student contacts, and training administrators to assure support for embedded assessment activities.

Redefining the Responsibilities of Teaching

Some teachers resist the role of assessor. Defining their work in terms of "delivering curriculum," these teachers speak of being responsible to "present material" and leave it up to the students to understand it. They don't think of themselves as responsible for analyzing student work. Said one teacher about use of an assessment strategy which relies heavily on teacher observations, *"I don't have time to observe my students. I am too busy teaching!"*

Similarly, some teachers appear to value commonly articulated standards for student performance but seem to see the exercise of working toward consensus on these as a threat to their autonomy. Asked to work with other teachers to assess student work up against standards described on a teacher-developed assessment scoring guide, one teacher remarked, *"I want to have time to create my own assessments, based on what I value."*

Other teachers expressed a mistrust of the assumption that teachers could come to share common performance standards: *"The other major issue is teachers scoring these assessments. Many do not trust their peers to score papers in an appropriate manner."*

Potential costs of implementing an instructionally sound assessment system include those related to redefining the responsibilities of teachers.

Changing Parent and Community Expectations

Teachers in alternative assessment projects recognize the fragile nature of their innovations. Parents and other community members expect the grades and test scores with which they are familiar. The implementation of instructionally sound assessments imply potential costs related to parent and community education. Said one teacher, *"The greatest challenge is educating parents so they can fully appreciate the data collected."*

Envisioning a New Paradigm

The changes demanded to implement instructionally sound assessments are profound. They require the individuals involved not only to learn new things, but to challenge long-held beliefs and to imagine things they have never seen before. Work in CAC projects reveals that many educators and parents still cling to skill-based instructional strategies, and the belief that many students are limited in what they can learn by innate ability. Teachers commonly spoke about the reluctance to change:

> *As teachers we are isolated. New teachers are getting different instruction and we teach the way we were trained. If there is no time for retraining we stay in the old ways and with what works for us over time. It's hard to give up the old ways.*
>
> *Some teachers are insecure in their own ability to pull in necessary subskills as they teach from a whole language basis. They want definite teaching guides, questions and answers, and hard data (e.g., % scores) to assign "grades." These teachers will never be successful with alternative approaches because they themselves are not "thinkers."*

Participants in innovative assessment projects worry that their work should be practical enough to be useful in current school structures but revolutionary enough to inspire change. They wonder if assessments should assess what most teachers now teach or those things we hope students will learn:

> *We need more models or exemplars of alternative assessment items or projects. It is very difficult to break the mold and have people really be creative.*
>
> *I think that school districts should develop a "sample" portfolio at each grade level. Individual sites could then modify or keep the sample portfolio. I think they should also develop sample assessment items for every grade level (a minimum of one per grading period) so that all children can get exposed to the new assessment - as well as all the teachers, including those reluctant to change.*

In detailing a framework for the costs of implementing an instructionally sound assessment system we must include the potential costs of supporting and pacing innovative change.

The investment of dollars, time, and energy required to assure that performance assessment actually improves student performance are high. Although many costs are associated with initial development work, many relate to the sustaining structures and processes which will assure that assessments continue to have a positive impact on teaching and learning. Yet some of the issues listed in this section as costs may prove to require no additional expense. Many of the costs of implementing an instructionally sound assessment system are measured in the intellectual and imaginative energy it will take to look at old structures, roles, and responsibilities in new ways.

A Framework for Analysis of the Benefits and Costs of Performance Assessment

Benefits of Performance Assessment

Enhancing Opportunities to Learn

- Increased Use of Performance-Based Instruction
- Increased Instructional Opportunities for Special Needs Students
- Increased Sensitivity to Student Diversity
- Increased Expectations for Student Performance

Enhancing Teachers' Content Knowledge and Expertise

- Improved Ability to be Explicit about Standards
- Alignment of Professional Development with Student Performance
- The Creation of Cultures of Collegial Inquiry
- Increased Teacher Consensus on Standards
- Creating Structures for Collegial Work
- Encourages Conscious Instructional Decision-Making
- Support for Teacher as Facilitator
- Increased Knowledge of Assessment

Improving Communication with Parents

- Improved Quality of Information on Student Performance
- Encourages Home Support of Student Learning

Guiding School Restructuring and Change

- Helps Sustain Restructuring
- Motivates Restructuring

Enhancing Student Achievement

- Increased Engagement and Motivation Toward Learning
- Increased Ability to Self Assess and Take Responsibility for Learning

Costs of Performance Assessment

Expenditures

- Personnel for Task Design and Development
- Teacher Orientation
- Personnel Time for Administration, Interpretation, and Scoring
- Project Facilitators
- Clerical Support
- Consultants
- Reporting Achievement Data
- Materials for Development, Administration, and Reporting
- Supplies Used in Assessments or Preparations

Challenges which Imply Costs

- Addressing Inequities in Opportunities to Learn
- Building Teacher Content Knowledge and Instructional Expertise
- Overcoming Cultures of Isolation
- Redefining Responsibilities of Teaching
- Changing Parent and Community Expectations
- Envisioning a New Paradigm

Putting the Cost/Benefit Frameworks to Use

It is clear that more data are necessary to accurately compare the relative costs and benefits of various assessment strategies. The frameworks provided here are designed to expand the dimensions on which we define "cost" or "benefit" and to give guidance to further research. For example, in order to determine relative costs/benefits, the next phase of analysis will require examination of the contrasting dimensions of these frameworks within individual projects.

At this time, few conclusions can be drawn. Although a large investment of dollars, time, and energy will continue to be required in order to develop performance assessments, there is great optimism that their use will have a positive impact on teaching and learning.

For now, the cost/benefit frameworks drawn from the initial phase of work in CAC projects suggest three important questions:

Are we under-estimating the costs of traditional assessments?

Standardized and text-based tests are commonly thought to provide a relatively inexpensive means of assessing student achievement. If we consider the value of the instructional activities these tests displace, and the potential expenditures to compensate for poorly prepared citizens and workers, we will need to revise cost estimates of traditional assessments upward.

Are we under-estimating the costs of assessments designed to improve teaching and learning?

Many of the estimates of the costs related to developing performance assessments refer to expenditures for task development and scoring processes. But performance assessments themselves, no matter how carefully designed to require high level content, no matter how authentic or reliably scored, will not guarantee improved teaching and learning. The estimates of costs related to instructionally sound assessments require consideration of building and sustaining the capacity of teachers and students to use assessment to improve their work.

Will the benefits of instructionally sound assessments off-set costs?

Although all the cost and benefit data on performance assessments are tentative, it must be noted that much of the beneficial impact observed or noted by participants in performance assessment projects provide direct counterpoint to many of the challenges which imply possible costs.

Although CAC data suggest that one of the major challenges of implementing instructionally sound assessments will be in rectifying inequities in opportunities to learn, there are also data which suggest that participation in performance assessment development efforts increases the use of performance-based instruction, enhances instructional opportunities for special needs students, increases teacher sensitivity to student diversity, and helps raise teachers' expectations for student performance.

In counterpoint to the costs inherent in increasing teachers' content knowledge and expertise is the observation that participation in performance assessment development efforts can positively influence teachers' understanding of the curriculum content and inspire greater alignment between professional development opportunities and students' instructional needs.

Although the implementation of instructionally sound assessments suggests costs related to redefining the responsibilities of teachers and increasing time for collaborative work, performance assessment development efforts have the potential to increase teacher consensus on standards, create structures for collegial work, encourage conscious instructional decision-making, and help teachers to become better facilitators and observers of student work.

And finally, although the education and support of parents, teachers, and policy-makers was considered a major potential cost related to the implementation of instructionally sound assessment, CAC projects point to the potential of the work to assuage these costs by improving the quality of assessment data available to parents and the community, improving students' ability to take responsibility for their learning, and motivating schools to sustain restructuring efforts.

CHAPTER 5

CONCLUSIONS AND IMPLICATIONS

CONCLUSIONS AND IMPLICATIONS

California is in the forefront of education reform. The curriculum frameworks and school level task force reports weave together professionally articulated curriculum content standards and the latest cognitive research to build a vision of schools in which all students learn to think critically, communicate effectively, and work cooperatively in teams. Essential to that vision is an assessment system which stimulates and supports meaningful teaching and learning.

In Senate Bill 662, California makes a commitment to develop a new statewide student assessment system "which has as its primary purpose, the improvement of instruction for pupils." Although established independently of the new statewide assessment system, the California Assessment Collaborative makes an important contribution to that effort by supporting and studying school and district assessment development and implementation efforts. As the result of that work, the CAC has, through this report, contributed

- a conceptual map detailing the requisite elements of an instructionally sound assessment system,
- examples and strategies which can be used by districts and schools considering their own assessment development work, and
- a framework for considering the potential costs and benefits of site-based assessment development and implementation projects.

It is hoped that these contributions assist schools to evaluate the potential of new assessment strategies to improve teaching and learning, plan their own local assessment development or implementation efforts, and consider ways to coordinate those efforts with the statewide system.

The fully realized vision of an instructionally sound assessment system brings to mind schools unlike most that we know today. In schools in which such a system were embedded, we would expect to find that all students were familiar with the standards for high quality performance and were offered the opportunities they needed to reach those standards. Students would be active participants in the learning process and would routinely assess their own work and progress toward the standards. Classroom instruction and assessment activities would require students to demonstrate what they know and can do through meaningful and engaging performance tasks. Teachers would base instructional planning on rich curricula designed to address a wide range of knowledge, skills, and capacities, and on their analysis of student work and progress toward performance standards.

Held up against this vision, schools in CAC projects each have a good ways to go. Yet when considered collectively, the preliminary development work and research on potential impact offers a set of conclusions and implications to guide those responsible for the support, development, and implementation of performance assessment systems.

1. The implementation of instructionally sound assessments will require a fundamental redefinition of what it means to teach.

Traditionally, the nature of teachers' work is described in the rhetoric of "coverage." Teachers speak of being required to "cover material," "get through the textbook," or "deliver the curriculum." Periodically along that road the teacher gives and grades tests designed to look back on what has been learned.

The implementation of instructionally sound assessments will require a fundamental change in the definition and practice of teaching. Teaching must entail not only the ability to plan and carry-out effective instructional activities, but the ability to assess student work accurately and fairly, and to use the data from that assessment to assure that all students have adequate opportunities to reach high standards of performance. Teachers must be seen, and must see themselves, as decision makers responsible for planning and carrying out an on-going cycle of instruction and assessment based on the demonstrated needs of individual students. Assessment must become the teachers' tool, a skill used to determine what students can do successfully, and a guide for planning future lessons, observations, or experiences.

Agencies which provide pre-service and in-service professional development programs, and school leaders and policy makers, must look to rectify the inconsistencies between this definition and current practice.

For all classroom teachers this means

- participating in collaborative efforts to reach common understandings of standards
- aiming instruction at the achievement of standards
- basing instructional planning on what students know and need to learn
- honing skills related to the analysis of student work

For leaders of curriculum, instruction, and assessment reform efforts this means

- providing participants access to state curriculum frameworks and school level task force reports in order to facilitate alignment with these documents
- structuring opportunities for teachers to work toward common understandings of standards
- structuring opportunities for all teachers to enhance their skills of analysis of student work
- working with parents and community groups to redefine their expectations for teachers' work

For district and state level policy makers this means

- acknowledging the centrality of teachers in any assessment system designed to improve what students know and can do
- providing opportunities for practicing teachers to build their capacity to use assessment to improve instruction
- providing leadership to educate the public and school community about the critical role of teachers as assessors
- providing leadership to the design of teacher pre-service programs to emphasize the critical role of teachers as assessors

2. The development and implementation of instructionally sound assessments must be understood as a complex, systemic process.

Work in CAC projects suggests that in order to assure that performance assessments improve teaching and learning, assessment development must go beyond the articulation of standards and invention of tasks and scoring procedures. It must be seen as the challenge of developing a system of interdependent dimensions which includes, in addition to the characteristics of the assessments themselves, attention to structuring the consequences of assessment so that they contribute to the capacity of teachers and students to improve their work. Work on each element of the system influences all other parts.

Seen as a multi-dimensional set of elements and influences, assessment reform is best addressed through multiple entry points. Schools committed to using assessments to improve teaching and learning can ill afford to focus on task design to the exclusion of capacity building issues or to build scoring and reporting formats without regard for the consequences of the data they report. When current school structures or practices inhibit the ability of teachers or students to do their best work, they must be reconsidered and revised.

The progress of assessment development work must be guided, not by an arbitrary linear progression, but by the degree to which the work on any element reflects or guides the best instructional practice and is owned and understood by those who must make use of it. "Next steps" must be determined by the assessed needs of the administrators, teachers, parents, and students who are the audience for assessment data.

For all classroom teachers this means

- integrating what they learn about assessment into an emerging model of support for curriculum reform and school restructuring

For leaders of curriculum, instruction, and assessment reform efforts this means

- planning efforts with an awareness of all necessary dimensions and elements
- monitoring the progress of development work
- recognizing that a single weak element may affect strength of an interrelated one

For district and state level policy makers this means

- providing leadership to coordinate reform initiatives
- providing leadership to community education efforts to build support for assessment reform

3. Instructionally sound assessments must root and grow in cultures of collegial inquiry.

Derived from the Latin "to sit by (as with a judge)," the word "assessment" describes a cognitive process of working with data to draw conclusions and make judgments. Assessments become instructionally sound when structured and used in such a way to assist educators and students to improve their work. Educational leaders as well must be given opportunities to participate in, and then be held accountable for, disciplined inquiry.

The "Five Year Cost and Implementation Plan" for the California Learning Assessment System recognizes that, in order for the new system to improve classroom instruction and student learning, it must be supported by at least two levels of professional development. Teachers must have opportunities to understand and internalize student performance standards by scoring student work, and in addition, they must have opportunities to develop and learn to use instructional strategies which will help all students reach those standards. That report anticipates that the California Department of Education will take responsibility for providing scoring opportunities and calls on the California Subject Matter Projects, SB 1882 Regional Staff Development Consortia, county offices of education, and post-secondary institutions to provide "much richer and deeper staff development that begins with assessment, but goes on to address the heart of teaching and learning."

The experience of CAC projects suggests that those professional development efforts will only bear fruit if the school systems they serve are themselves learning organizations which support cultures of collegial inquiry. When teachers in CAC projects spoke of the potential of performance assessment to improve teaching and learning, they most often related that potential, not to having a new assessment tool, but to being themselves engaged in collaborative action-research projects in which they operationalized standards, constructed and revised tasks, interpreted student work, and analyzed what it would take to help students perform better. Some of their greatest doubts and frustrations came from feeling that project work was isolated from the everyday context of school life. Because assessment is an integral part of the teaching and learning process, action-research of this nature must be embedded into the culture and calendar of schools and school systems.

Implementing a system of instructionally sound assessments will require embedding into school practice structures in which administrators and teachers regularly have opportunities to collaborate on analyses of student performance, to conduct investigations on questions of common concern, and to determine the implications of the results for instructional and program planning. The norms of schools and school systems must emphasize disciplined inquiry and collaboration.

For all classroom teachers this means

- collaborating with colleagues and curriculum experts through the course of development and implementation work
- honing skills of observation and analysis of student work
- being both rigorous and patient through various reviews and revisions necessary in assessment development work

- working with all students to develop the skills of self-assessment

For leaders of curriculum, instruction and assessment reform efforts this means

- structuring the development process to include collaborations, multiple revisions, sustained effort and quality feedback
- providing access to facilitators and coaches for teachers involved in development or implementation efforts
- coordinating curriculum, assessment, and professional development efforts
- involving all teachers in opportunities to analyze tasks and consider appropriate instructional strategies, especially when assessments have been externally produced
- preparing all teachers to be critical consumers of assessments

For district and state level policy makers this means

- empowering teachers as central decision makers in student learning
- providing leadership toward building school cultures in which the analysis of student work is valued and expected
- coordinating existing professional development networks around the effort to know and operationalize standards, building teacher capacity and student access and capacity
- providing leadership to see that professional development opportunities for teachers stress meaning-making and applications, not just delivery of information
- providing leadership and support to sustained, long term reform efforts

4. Equity in assessment requires attention to the development of fair and unbiased tasks and to equitable opportunities to learn.

In addition to eliminating bias in assessment tasks and scoring processes, the implementation of instructionally sound assessments will require that all students have equitable opportunities to reach the performance standards. Equity requires that all students, regardless of their ethnic, linguistic, social, or economic backgrounds, be given access to a meaning-centered, thinking curriculum which is relevant to their cultural orientation and life experiences. Instructional and assessment tasks must recognize differences in learning styles, interests, and language proficiency and should provide students opportunities to express themselves in more than one way. Although all students should be expected to reach the same high standards of performance, equity requires giving each student the time needed to master a given level of performance.

Educators at all levels must take responsibility for rigorous and on-going analysis of the opportunities available for all students to meet high standards and for the restructuring of school and instructional programs which may be necessary for them to do so.

For all classroom teachers this means

- examining expectations for student performance
- providing all students access to meaning-centered, thinking curricula and instruction
- providing students opportunities to demonstrate what they know through multiple modes of expression
- providing the time and other resources necessary for each student to perform well

For leaders of curriculum, instruction and assessment reform efforts this means

- assessing the access of all students to meaning-centered, thinking curricula and instruction

- conducting reviews of assessment and instructional materials to see that they are free of bias and permit multiple modes of expression

- structuring schools to increase the capacity of both teachers and students to improve their work

For district and state level policy makers this means

- determining, monitoring, and promoting policies which include attention to equity and access

- providing leadership to restructuring efforts which increase the capacity of teachers and students to improve their work.

5. Assessment development is not just about doing things right, it's about doing the right things.

There is a danger in times of shrinking resources to think of assessment reform as simply a costly innovation, "one more thing" to challenge a profession already stretched thin. Assessment development work is often done in isolation from the context of school life and exists at the behest of special program funds or grants, vulnerable to the whims of budget makers or competition from the next "innovation of the year."

But the real challenge of assessment reform is not about what to put in a portfolio, how to score it, or even what professional development is necessary to help teachers use it effectively. Assessment development and implementation is a vehicle for exploring questions at the very heart of the purposes and processes of schooling:

> *What do we want students to know and be able to do?*
>
> *How will we know they can do it?*
>
> *What resources must be available to all students to assure that they succeed?*
>
> *How do we structure and pace the instructional program to prepare all students to succeed?*
>
> *What are the appropriate roles for teachers, administrators, parents and policy makers in assuring appropriate opportunities for all students?*
>
> *What does it mean to be "accountable"?*

These are questions central to the work of every classroom teacher, school administrator, and education policy maker, and they must be answered by those most directly responsible for making schools work. They will not be answered by the ad-hoc "Alternative Assessment Task Force" or by implanting the latest testmaker's product into schools as we know them today.

Site level alternative assessment development efforts are at their most valuable as a heuristic process for school restructuring: when their work is defined, not as making performance assessments, but as inciting all those connected with schools today to examine, and as necessary, restructure the work they do.

For all classroom teachers this means

- taking up the responsibility to be a fair and accurate assessor

- beginning now at least to experiment with new assessment strategies

For leaders of curriculum, instruction and assessment reform efforts this means

- involving *all* teachers in exploring the role of assessment in instruction

For district and state level policy makers this means

- recognizing that teacher centrality in assessment development and implementation is the only way to assure increased student performance

- linking assessment development and implementation and restructuring efforts

- providing continued support for district and site-based performance assessment development and implementation efforts.

REFERENCES

REFERENCES

California Department of Education. (1992). *SB 662 A New Statewide Student Assessment System: Five Year Cost and Implementation Plan. A Report to the Governor and Legislature.* Sacramento, CA: Author.

California High School Task Force. (1992). *Second to None: A Vision of the New California High School.* Sacramento, CA: California Department of Education.

California Learning Assessment System Policy Committee. (1993). Some Principles and Beliefs about the Role of Assessment in California's School Reform Plan. As reported in *California Curriculum News Report*, 18 (4).

Congress of the United States Office of Technology Assessment. (1992). *Testing in American Schools: Asking the Right Questions.* Washington, DC: Author.

Darling-Hammond, L. (1991). The Implications of Testing Policy for Educational Quality and Equality. *Phi Delta Kappan*, 73 (3), 220-225.

———. (1990) Instructional Policy into Practice: 'The Power of the Bottom over the Top.' *Educational Evaluation and Policy Analysis*, 12 (3), 339-347.

———. (1992). *Standards of Practice for Learner-Centered Schools.* New York: National Center for Restructuring Education, Schools, and Teaching.

Elam, S.M., Rose, L.C., & Gallup, A.M. (1992) The 24th Annual Gallup/Phi Delta Kappa Poll of the Public's Attitudes Toward the Public Schools. *Phi Delta Kappan*, 74 (1), 41-53.

Elementary Grades Task Force. (1992). *It's Elementary!* Sacramento, CA: California Department of Education.

Fleming, M. & Chambers, B. (1983). Teacher-Made Tests: Windows on the Classroom. In W. E. Hathaway, ed., *Testing in the Schools. New Directions for Testing and Measurement.* San Francisco: Jossey-Bass, 19:29-38.

Frederiksen, J.R. & Collins, A. (1989). A Systems Approach to Educational Testing. *Educational Researcher*, 18 (9), 27-32.

Goodman, K.S., Bird, L.B., & Goodman, Y.M. (1989). *The Whole Language Evaluation Book.* Portsmouth, NH: Heinemann Educational Books.

Greenleaf, C. (1992). The Impact of the HERALD Project on Student Learning. *HERALD Project Evaluation Report.*

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991). Complex, Performance-Based Assessment: Expectations and Validation Criteria. *Educational Researcher*, 20 (8), 15-21.

Madaus, G.F., West, M.M., Harmon, M.C., Lomax, R.G., & Viator, K.A. (1992). *The Influence of Testing on Teaching Math and Science in Grades 4-12.* Executive Summary of a study funded by the National Science Foundation (SPA8954759) and conducted by the Center for the Study of Testing, Evaluation, and Educational Policy, Boston College.

Middle Grade Task Force. (1987). *Caught in the Middle: Educational Reform for Young Adolescents in California Public Schools.* Sacramento, CA: California Department of Education.

Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools.* New York: The Free Press.

Mt. Diablo Unified School District. (1992). *Curriculum-Based Alternative Assessment for Mathematics: A Summary of 1992 Results.* Mount Diablo, CA: Author.

New Standards Project. Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA; National Center on Education and the Economy, Rochester, NY: Author.

O'Neil, J. (1993). On the New Standards Project: A Conversation with Lauren Resnick and Warren Simmons. *Educational Leadership*, 50 (5), 17-21.

Redwood City School District. (1992). *Assessment Sampler*. Internal staff newsletter.

Resnick, L.B. & Klopfer, L.E. (1989). Toward the Thinking Curriculum: An Overview. In Resnick, L.B. & Klopfer, L.E., eds., *Toward the Thinking Curriculum: Current Cognitive Research* (1989 Yearbook of the Association for Supervision and Curriculum Development). Alexandria, VA: Association for Supervision and Curriculum Development, 1-18.

Shulman, L. S. (1980). Test Design: A View from Practice. In E. L. Baker and E. S. Quellmalz, eds., *Educational Testing and Evaluation*. Los Angeles: Sage, 63-73.

Simmons, W. & Resnick, L. (1993). Assessment as the Catalyst of School Reform. *Educational Leadership*, 50 (5), 11-15.

Snyder, J. (1992). *A Community Accountable: Values and Standards at CPEI*. A paper presented at the annual meeting of the American Educational Research Association.

Stiggins, R.J. & Conklin, N.F. (1992). *In Teachers' Hands: Investigating the Practices of Classroom Assessment*. Albany, NY: State University of New York Press.

Wigdor, A.K. & Garner, W.R., eds. (1982). *Ability Testing: Uses, Consequences, and Controversies*. Washington, DC: National Academy Press.

Wiggins, G. (1991). Standards, Not Standardization: Evoking Quality Student Work. *Educational Leadership*, 48 (5), 18-25.

________. (1989). Teaching to the (Authentic) Test. *Educational Leadership*, 46 (7), 41-47.

Wiggins, G., Browne, J., & Houston, H. (1991). *Standards, Not Standardization* (Video and Print Curriculum). Stow, MA: Greater Insights Productions.

Williams, S. *Survey of Countywide Testing And Assessment Practices —1992-93*. San Diego: San Diego County Office of Education.

APPENDICES

A: RESEARCH METHODOLOGY

B: PROJECT CONTACTS

C: CREDITS

APPENDIX A: RESEARCH METHODOLOGY

Introduction

As stated in the legislation of AB 40, the purpose of the Alternative Assessment Pilot Project is to develop, field test, and validate a variety of alternative assessment instruments; conduct research and analysis of assessment practices to provide practitioners and policy makers with information on viable alternative strategies; and disseminate information on alternative assessment to a variety of stakeholders. These goals resulted in two major strands of research:

- to document and describe the *processes* for developing alternative assessments used in 22 pilot projects, and
- to describe the *effects* of developing and implementing alternative assessments on teachers, students, schools and districts.

This appendix to ***Charting the Course Toward Instructionally Sound Assessment*** provides readers with a description of our research methodology designed to gather and understand data related to these two strands.

Initially, the approach to understanding the *process* of developing alternative assessments focused on such issues as

- the impetus for change,
- ties to existing assessment practices,
- steps for task development,
- scoring procedures,
- standards setting,
- the characteristics of staff involved,
- reporting procedures, and
- the relationship to restructuring.

Our observations of project work sessions and analyses of project artifacts revealed that the development process was not a series of sequential steps but a recursive cycle of linked decision points and practices. Our analysis of the development process in pilot projects yielded a conceptual framework of key dimensions and elements in assessments designed to improve teaching and learning.

Our research effort also focused on describing the *effects* of developing and using alternative assessments on students, teachers, schools and districts. We documented both positive and negative effects by examining the CAC participants' perceptions of the costs or challenges and the benefits of their involvement. We also collected some limited information from each project about actual expenditures and resources required to carry out their development work.

The Research Team

The CAC is a unique research and development effort that provides technical assistance to projects and documents their processes and effects. The CAC staff consists of a research team from Far West Laboratory for Educational Research and Development (FWL), and coordinators and teacher facilitators in each of the member consortia of school districts. All staff members supplied data for our research.

The FWL staff was predominately responsible for designing research instrumentation and facilitating data collection procedures. The Consortia and FWL staffs worked together to provide technical assistance to projects in formal and informal ways. Frequently, staff attended planning, development, training and scoring meetings at which they acted as participant observers or provided technical support. Occasionally, staff were asked to facilitate meetings or present on a requested topic.

Consortia staff were assigned to specific pilot projects to assist with development whenever possible and attend meetings as a participant observer if no technical assistance was requested. FWL staff attended at least one development meeting as a participant observer and/or visited classrooms for all but two pilot projects.

Data Collection and Analysis Procedures

Both qualitative and quantitative data collection procedures were used to document and describe the *process* and *effects* of developing and implementing alternative assessments. Each procedure is described below.

Alternative Assessment Action Plans

Early during our first year of work (Fall 1991), each CAC project was asked to complete an Alternative Assessment Action Plan (AAAP). The AAAP was designed to capture a project's early history and process of development. (The AAAP form is available in paper or electronic form by request through FWL.) The major categories of information described in the AAAP included:

- characteristics of schools and personnel involved;
- curriculum areas covered;
- description of the assessment tasks, instruments or models being developed;
- predicted outcomes including expected impact on teachers, students, and schools;
- anticipated products including assessment tasks, scoring rubrics, training materials, etc.;
- planned development activities for the 1991-92 school year;
- a self evaluation plan.

An additional section asked for a detailed description of the necessary resources and costs for developing and implementing the projects. The form asked pilot leaders to describe the prerequisite knowledge necessary for teachers to participate, the kinds of technical support needed during the CAC contract period, and the pilot expenditures for human resources and materials both prior to and during the CAC funding period. Projects were also asked to report the funding sources in addition to CAC that supported their work.

Prior to writing their AAAP, project leaders met individually with FWL staff to review the AAAP form, talk through possible entries, and clarify the level of detail requested. During this meeting, project leaders were supplied with electronic disk versions of the AAAP as a template that could be expanded or reduced as needed. To reduce the data collection burden, leaders were also encouraged to use information from other project documents, to cut and paste text electronically, or to refer to existing documents which they were instructed to attach. The completed forms and their attachments provided the CAC with an elaborate description of the projects in a format that helped to guide technical assistance planning. Several project leaders who completed the forms remarked that spending the time talking with their colleagues and writing about their ideas in the AAAP helped them clarify project goals and intended timelines.

Field notes

While working with projects at formal meetings or site visits, staff recorded the projects' development processes and effects in field notes. These field notes documented the key issues discussed, problems raised, teacher anecdotes or strategies shared, and/or major decisions made. Staff notes were more than just a documentation of what happened in a meeting or site visit — they also reflected interpretations of these events as they revealed the projects' process of development and/or its impact on participants. Staff recorded suggestions for what the project might want to consider next, or highlighted what they perceived as potential obstacles or successful strategies.

Field notes were also gathered during occasional classroom visits to capture the actual use of the assessments being developed and characteristics of the context necessary to implement them; as well as during informal conversations with teachers about their work and its impact on their students.

The field notes were stored in individual project files at FWL and were occasionally copied and shared among all CAC staff. Staff meetings were used to discuss observations in the field. These discussions often led to technical assistance strategies or records of problems or successes the projects experienced. Field notes also provided a valuable archive of anecdotes which contributed to our understanding of project effects.

Project Artifacts

Throughout our first phase of work, staff collected assessment samples, meeting agendas and other artifacts. These artifacts, along with field notes taken at project sites and the AAAP's, were compiled for each project. To augment these collections, in the spring of 1992 each project was asked to submit additional artifacts which would help to document their development process and the current status of their work.

Projects contributed successive drafts of standards documents, assessment tasks, and scoring rubrics. Other common sources of evidence for the process of development included meeting or workshop agendas, minutes; and anecdotes from planning, development or scoring sessions. Sometimes, though rarely, artifacts included student work samples (occasionally scored or with teacher feedback recorded), completed student self evaluation forms, intact student portfolios, audiotapes and videotapes of student exhibitions of their work or conferences with teachers. Other less common contributions included reports of student assessment results, research articles, technical reports, and newsletters that described the projects or disseminated their findings. A few projects submitted binders containing training materials or manuals for administering the assessments.

Staff worked with project leaders to compile the artifacts. In some cases informal meetings were scheduled so that the project leader could discuss each artifact and explain its relevance to the project's development. These artifacts and project leader descriptions became a crucial data source for the CAC to understand each project's progress, as well as to identify the common challenges and stumbling blocks across all projects. Some project leaders reported that the collections served as celebrations of successes to date and reminders of the hard work ahead.

Focus Group Interviews

From January to March 1992, FWL staff conducted focus group interviews in each project to assess the effects of teachers' participation in developing and using alternative assessments. The focus groups were structured around three broad categories of questions:

- What have been the effects on teachers?
- What have been the effects on students?
- What have been the effects on the school or district?

The specific interview questions encouraged teachers to describe both positive and negative impact in each of these areas.

The questions directed at "effects on teachers" addressed the influence of teacher participation in assessment development, on curriculum and instructional practices, professional development, and work with colleagues. Questions related to "effects on students" focused on achievement, attitudes and self-evaluation. The "effects on school or district" were captured with questions on grading, accountability, and restructuring. Finally, teachers were also asked about their perceptions of the greatest costs or challenges, as well as the benefits of their involvement in assessment development and implementation. Specific interview questions are included in the Interview Protocol which is available from Far West Laboratory.

The interviews were scheduled to last about one hour, but constraints or flexibility in teacher time led to a range of 45 minutes to almost two hours. Most interviews lasted about one hour and fifteen minutes.

More than one interview was carried out in projects where it was clear that the implementation or development efforts might be experienced differently by participants from different contexts. Three different groups of teachers from the ACE K-2 Project participated in focus groups. One group represented veteran teachers from the initial development team, a second included an entire school level team of implementers, and the third represented teachers from the second phase development team. The San Francisco Rockefeller project had two focus group interviews — one from each of the participating schools. Finally, the La Mesa/Spring Valley portfolio project had two groups, one comprised of elementary school teachers and the other of middle school teachers. A total of 26 focus group interviews were completed with 108 teachers participating.

The interviews were conducted in a relaxed atmosphere, in which participants were encouraged to express their views candidly. Though a common protocol of questions led all focus groups to address the same issues, our follow-up prompts for examples or elaborations allowed us to go beyond surface comments and tailor the interviews to the interests and experiences of the groups of teachers who participated. Many teachers remarked that the focus group was a pleasant opportunity to discuss their project with colleagues and a welcome avenue for learning how others had implemented ideas in their classrooms.

For most interviews, two FWL staff were present. One staff member facilitated the interview, asked questions, and recorded notes with pen and paper, and a second used a notebook computer to transcribe teachers' comments. The interviews were also audiotaped. The comments contained in both sets of interview notes were summarized in one page charts organized according to the major categories of questions (a sample summary is available from Far West Laboratory). These summaries were later coded and tallied across projects to highlight anecdotes and examples of the most common benefits, costs, and challenges. The same coding scheme, briefly described below, was used for the open-ended questions on the year-end questionnaire.

A total of 1037 comments (817 positive impact and 220 negative) were coded. The most frequently occurring comments during the focus group interviews were related to positive effects on students and changes in curriculum and instructional practices.

Survey of Participants

Late in the spring of 1992, each participant in a CAC project was asked to complete a questionnaire (also available from Far West Laboratory) that explored the extent to which issues in the focus group interviews were perceived in the same way by a larger sample of teachers. In addition to the questions matched to the focus group topics of effects on teachers, students and schools, the questionnaire asked participants to compare some specific challenges with the benefits associated with alternative assessment design and use; and to contrast the usefulness of traditional and alternative assessments for a variety of purposes.

Most of the questions asked participants to respond on a five or seven point Likert-type scale. Several open-ended questions allowed teachers and administrators to provide their individual views on the greatest challenges and benefits of their participation, as well as any additional comments, reactions, or elaborations.

Two hundred and eighty-eight questionnaires were completed, representing a return rate that varied widely by project — one small project didn't return any surveys and nine projects of varying size had a return rate of virtually 100%. More than three quarters of the respondents were regular classroom teachers, with the remaining respondents made up of resource teachers or specialists, project or department heads, and a few school psychologists and principals.

The mean teaching experience of respondents was fourteen years (minimum = 0.75 years; maximum = 39.00 years). On average, teachers had been associated with their assessment projects for less than two years ($\overline{x}$ =1.67 years) with a range of a few months to no more than four years.

Knowing that participants' roles within the assessment projects also varied, we asked pilot leaders to indicate the level of involvement for each teacher in their project in a box on the top of the questionnaire. This was done prior to sending the forms to teachers. Pilot leaders were to record a "1" if the teacher served as an assessment "developer." "Developers" were integrally responsible for the design of tasks, portfolio templates, and scoring and interpretation schemes; and typically, the first to try out assessments and suggest revisions. A "2" indicated that the teacher served as a "field tester." "Field testers" were responsible for trying out an early version of an assessment in their classroom for the purpose of providing feedback to the developers. Some of these field testers joined the development teams as the assessments were refined for future use.

The final group of participants (indicated by a "3") represented assessment "implementers." These teachers typically had little contact with assessment developers and administered the assessment to their children at the request of district or project leaders. Very few of the CAC's projects were yet using their new assessments in classrooms beyond those of developers, so most of the "implementers" in our sample came from a single pilot project (Mt.Diablo Unified). When the questionnaires were returned, the sample comprised 30% developers, 19% field testers, 31% implementers, and 20% unknown. Our initial analyses revealed very few differences in the response patterns of each type of participant. Because there were a large number of respondents whose role was not known, we aggregated the results across all groups.

Frequencies of participants' responses to each question by level of response on the Likert-type scales were calculated. Participants' responses to the open-ended questions on the greatest challenges and benefits of developing and using alternative assessments and any further comments or elaborations were coded and tallied. The coding scheme was based on three general categories (effects on teachers, students, and school/district). Items from the questionnaire were used as initial codes, with additional codes added for comments not captured by existing items. Comments had both negative and positive connotations and were coded accordingly. The total number of coded comments was 531 for benefits and 438 for challenges.

The focus group interviews and questionnaire revealed that teachers and administrators were overwhelmingly positive about the effects of alternative assessment development projects on teachers and students, but these data represent respondents' perceptions of the effects for projects that are, in most cases, still considered works-in-progress. To date, we have limited data to verify whether perceived positive effects can be attributed to the assessment strategy itself, or the degree to which the perception of benefit is supported by actual student or teacher performance. The findings from the survey and interviews presented in the body of this report have been tempered at times by contradictory evidence from other data sources.

Putting It all Together

The research conducted by the CAC is by nature exploratory. At this early stage in what may be a revolution in assessment practice, it is important that before setting up formal studies, we are sure we are asking the right questions. The constant tension in our work has been, and will continue to be, how to balance the need for validated assessment models with the realization that instructionally sound assessment is more a dynamic strategy than a static tool. In this report we have drawn on all our data sources both to illustrate that case, and to consider its implications for future development and implementation efforts, analysis of costs and benefits, policy development and further research.

Conceptual Map

The conceptual map is a composite of the collective experience of all CAC projects, drawn from the bird's-eye vantage point of CAC staff. The elements are practices observed in one or more projects which influenced the potential of the assessment strategy to impact teaching and learning. The dimensions are categories of those elements. Together they serve as our hypothesis about the dimensions and elements necessary to an instructionally sound assessment system.

We came to this hypothesis using data gathered primarily from observations of project development work and artifacts. These data also provided a source for illustrative examples of the ways which some elements might be successfully addressed. The map was drafted and circulated among CAC staff and several project leaders who contributed to numerous progressive revisions and suggested the illustrative examples. ***Challenges, Cautions, Caveats*** emerged from the discrepancies between our hypothesis for instructionally sound assessment and the realities of work in many projects. They represent often unattended issues or stumbling blocks common across several projects.

First Person Narratives

Having made the choice to look across 22 pilot projects with similar questions, we sacrificed the opportunity to go in-depth with any. Although the conceptual map and its illustrations provide a comprehensive hypothesis which we believe can guide future assessment development and implementation work, depth of understanding of the process or impact of any one project has been lost in the trade-off. These narratives are included in an attempt to amplify the voice of project participants and help readers understand how those participants talk about the evolution of their work. Although the narratives have been edited for readability, they have not been influenced by other data sources. They are solely the opinion of their authors.

Cost/Benefit Framework

The costs and benefits of traditional assessments were described in a recent study conducted by the Office of Technology Assessment of the U.S. Congress. The framework for analysis described in this report is drawn from review of this study, a study of testing programs in the San Diego County school districts and additional comments from school administrators involved in other pilot projects.

Data on the relative costs and benefits of alternative assessments were harder to compile. The CAC's pilot projects are each unique in size and scope, and as works-in-progress provide little opportunity as yet to calculate true costs or to attribute benefits. The framework for analysis of costs and benefits of alternative assessments was derived from the collective experience of all projects. The framework for costs describes both the expenses listed in project action plans and the costs suggested by the common challenges observed in development work and described in participant interviews and surveys. Data on potential benefits was compiled from surveys, interviews and observations of on-going development work.

Before being finalized, the framework was reviewed and revised by FWL and consortia staff and selected project and other school leaders.

Conclusions and Implications

One of the biggest challenges in the work of the California Assessment Collaborative is to draw from 22 diverse pilot projects, each with its own unique issues, conclusions which will inform school leaders, policymakers, and teachers throughout the state. In attempting to do so, we sought to identify "big ideas" which seemed to emerge across all pilot project development work, and in the data on project effects.

The conclusions stated in the close of this report represent themes which emerge from all of the data available. These themes were discussed among consortia and Far West Laboratory staff throughout the data collection and analysis process. As they emerged the writing committee identified commonalities and drafted an initial set of conclusion statements. Each conclusion was followed by a set of possible implications for the work of each of the key audiences for this report.

A draft of the *Conclusions and Implications* chapter was reviewed and revised by members of the CAC Governance Team (which includes representation from all partner groups within the Collaborative) after each had had the opportunity to read the penultimate draft of this report. Following that discussion, the chapter was submitted to all CAC staff and Governance Team members for final review.

Future Research

As described in this report, our research on the CAC's projects has perhaps revealed more questions than answers. We have suggested a conceptual map of the key dimensions and elements in developing and implementing an instructionally sound assessment system. For these dimensions to be fully realized, a great deal must be learned about processes and practices that allow them to work together systemically. Our continuing research must also examine the progress our pilot projects are making in implementing these dimensions and elements. This progress is defined not just in superficial attention to each, but in terms of the degree to which project work reflects or guides the best instructional practice, and is owned and understood by those who must make use of it.

We have suggested a framework for researching costs and benefits — re-examining the costs of traditional assessment, re-estimating what it costs to design and implement assessments that will improve teaching learning, and re-explicating the ways in which benefits may off-set costs.

This cost/benefit framework also informs our emerging research agenda. Our continuing research will consider examining costs and benefits in specific contexts rather than across all projects. Future project-specific research will take a deeper, more focused look at the degree to which reported, or potential, benefits have been experienced, and look at the specific design or implementation characteristics that made them possible.

APPENDIX B: CAC PILOT PROJECT CONTACTS

Greater ACE Consortium Projects	*Contacts*
ACE Consortium K-2 Project	Kate Jamentz Director California Assessment Collaborative
All Students Will Succeed	Mary Lairon Assistant Superintendent Redwood City School District
Curriculum-Based Assessment in Mathematics	Richard Nicoll Director, Research and Evaluation Mt. Diablo Unified School District
Curriculum-Based Performance Assessment **Curriculum-Based Testing in Special Education**	Catherine Gallegos Director of Personnel and Student Services San Lorenzo Valley Unified School District
Foreign Language Proficiency	Barbara Liddell Associate Superintendent, Educational Services Palo Alto Unified School District
HERALD Portfolio Project	Cindy Greenleaf HERALD Project San Francisco Unified School District
Technology-Based Writing Portfolios	Sheila McGann-Tiedt Principal, Stevens Creek School Cupertino Union School District
PACE — San Francisco	Jan Link Testing, Evaluation, & Research San Francisco Unified School District

Greater San Diego Area Consortium	
Chapter 1 Alternative Assessment Project	Dan Nasman Assistant Superintendent, Educational Services Lakeside Union Elementary School District
Elementary Portfolio Assessment Project	Cathy Young Curriculum Resource Teacher La Mesa/Spring Valley School Distrtict

End-of-Course Assessment Project	Bill Brock Director of Testing & Evaluation Grossmont Union High School District
Coordinator Science Portfolios	Vivienne Deutsch Teacher Monte Vista High School
GATE Identification Portfolio Project	Bruce Montgomery Coordinator, Curriculum and Instruction Oceanside Unified School District
LEP Student Assessment Project	Shirley Day Teacher, Black Mountain Middle School Poway Unified
Math/Science Comprehensive Assessment Project	Janet Abbott Coordinator, Mathematics and Science Chula Vista Elementary School District
San Diego City Schools	
Elementary Science and Language Arts Assessment Development **Hispanic Reading Project** **New Standards Project** **UCLA/CRESST Pilot in Social Studies**	Linda Carstens Specialist, Performance Assessment Division San Diego City Schools
PACE — San Diego	Karen Bachofer Specialist, Performance Assessment Division San Diego City Schools
California Learning Record	Mary Barr Director University of California, San Diego

APPENDIX C: CREDITS

Principle Author

Catherine Jamentz
Director
California Assessment Collaborative

Director of Research

Andrea Whittaker
Research Associate
Far West Laboratory for Educational
Research and Development

Writing Committee

Linda Carstens
Specialist, Performance Assessment Development Unit
San Diego City Schools

Janice Goertz,
Teacher-Facilitator
California Assessment Collaborative

Kendyll Stansbury
Senior Research Associate
Far West Laboratory for Educational
Research and Development

Sandra S. Williams
Director, Assessment and Planning,
San Diego County Office of Education

Carl Zon,
North Consortium Director
California Assessment Collaborative

Contributions

Karen Volz Bachofer
Program Evaluator, Performance Assessment
Development Unit
San Diego City Schools

Don Barfield,
Deputy Director
Far West Laboratory for Educational
Research and Development

Susan Bennett,
Consultant
California Learning Assessment System
California Department of Education

Fernanda Berman,
Administrative Assistant
California Assessment Collaborative

Ruben Carriedo
Assistant Superintendent
San Diego City Schools

John Cradler,
Director of Educational Technology
Far West Laboratory for Educational
Research and Development

Jerilee Fischer-Garza,
Resource Teacher, Performance Assessment
Development Unit
San Diego City Schools

Catherine Geanuracos,
Administrative Assistant
California Assessment Collaborative

Jack Gibbany,
Administrator, Curriculum Services
Palo Alto Unified School District

Carolyn Howard-McBride,
Teacher
Piedmont Avenue School
Oakland Unified School District

Amy Shively Jackson,
Research Associate
Far West Laboratory for Educational
Research and Development

Nanette Koelsch,
Research Associate
Far West Laboratory for Educational
Research and Development

Judith Lookabill,
Assistant Superintendent,
Instructional Services
Santa Clara County Office of Education

John Parsons,
Project Coordinator
California Assessment Collaborative
San Diego County Office of Education

Judith Swinko,
Resource Teacher, Performance Assessment
Development Unit
San Diego City Schools

Editor

Gerald Camp,
Administrative Assistant, The National Writing Project
University of California, Berkeley

NOTES